Modelling Approaches via
(Batch) Markov modulated Poisson processes

by

Yoel Gustavo Yera Mora

A dissertation submitted by in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in

Ph.D. Program in Mathematical Engineering

Universidad Carlos III de Madrid

Advisors:

Rosa E. Lillo
Pepa Ramírez-Cobo

Junio 2020

To my family,
for their endless support

Acknowledgements

When a page is closed in life, one usually looks back and tries to thank all the people who have contributed in one way or another to it. A doctoral thesis is one of those pages, a page that goes beyond four or five years, since to get to that page many others have had to be closed before. To name all the people who in one way or another have contributed to my formation and personal development is impossible, but to all of them I would like to show my gratitude.

Specially I would like to thank my advisors Dr. Rosa E. Lillo and Dr. Pepa Ramírez-Cobo, for their permanent support, encouragement, and guidance throughout my Ph.D. studies. Without you, this thesis would never have come up.

I would like to thank all my Ph.D. fellows, particularly to Verónica, Alba, Elisa and Juan Carlos, for their friendship and advice. Also, I want to extend my gratitude to my cuban fellows Miraine, Josue, Leonel, Alina, and Jose for their friendship from the distance.

I would like to thank also to my all university professors, who were strong pillars in my training. Especially to Dr. José E. Valés because thanks to him I discovered probability and statistics. During my studies in Cuba he was my guide and inspiration for the study of mathematics.

Finally, and most importantly, I wish to thank my wife, Claudia, for her constant support, love and affection. To my children Balaram and Anandi, for their kisses, hugs and smiles that made me draw strength from where there was none to go on. To my parents, for their love and constant support throughout every step of my life. In spite of the distance I know that you have always carried me in your innermost heart.

I also want to extend my gratitude to Fabrizio Ruggeri and Bo F. Nielsen, for their collaboration, suggestions and support in the last chapter.

I am grateful to the Department of Statistics of Universidad Carlos III de Madrid, for providing the financial support to carry out this thesis.

The work presented in this thesis was partially supported by research grants and projects MTM2015-65915-R and ECO2015-66593-P (Ministerio de Economía y Competitividad, Spain) and P11-FQM-7603, FQM-329 (Junta de Andalucía, Spain) and Fundación BBVA.

Published and submitted content

- Yoel G. Yera, Rosa E. Lillo and Pepa Ramírez-Cobo (2019) Findings about the BMMPP for modeling dependent and simultaneous data in reliability and queueing systems. *Applied Stochastic Models in Business and Industry*, 35(2):177-190. DOI: 10.1002/asmb.2327.

  This publication is entirely included in chapter 2 of the thesis.

- Yoel G. Yera, Rosa E. Lillo and Pepa Ramírez-Cobo (2019) Fitting procedure for the two-state Batch Markov modulated Poisson process. *European Journal of Operational Research*, 279(1):79-92. DOI: 10.1016/j.ejor.2019.04.018

  This publication is entirely included in chapter 3 of the thesis.

- Yoel G. Yera, Rosa E. Lillo, Pepa Ramírez-Cobo, Fabrizio Ruggeri and Bo F. Nielsen (2020) A bivariate two-state Markov modulated Poisson process for failure modeling.*Submitted for publication.*

  This publication is entirely included in chapter 4 of the thesis.

Abstract

This dissertation is mainly motivated by the problem of statistical modeling via a specific point process, namely, the (Batch) Markov Modulated Poisson process. Point processes arise in a wide range of situations in physics, biology, engineering, or economics. In general, the occurrence of events is defined depending on the context, but in many areas are defined by the occurrence of an event at a specific time. Sometimes, in order to simplify the models and obtain closed form expressions for the quantities of interest, the exponentiality and/or independence of the inter-event times is assumed. However, the independence and exponentiability assumptions become unrealistic and restrictive in practice, and therefore, there is a need of more realistic models to fit the data.

The Batch Markov Modulated Poisson Process (*BMMPP*) is a subclass of the versatile Batch Markovian Arrival process (*BMAP*) which has been widely used for the modeling of dependent and correlated simultaneous events (as arrivals, failures or risk events). Both *BMMPP* and *BMAP* allow for dependent and non-exponentially distributed inter-event times as well as for correlated batches, but at the same time they inherit the tractability of the Poisson processes. Therefore, they turn out suitable models to fit data with statistical features that differ form the classical Poisson assumptions. In spite of the large amount of works considering the *BMAP* subclasses of processes, still there are a number of open problems of interest that will be considered in this dissertation, which is organized as follows.

In Chapter 1, a brief theoretical background that introduces the most important concepts and properties that are needed to carry out our analyses is presented. The markovian point processes and their main properties are introduced.

In Chapter 2 the identifiability of the stationary $BMMPP_m(K)$ is proven, where K is the maximum batch size and m is the number of states of the underlying Markov chain. This is a powerful result for inferential issues. On the other hand, some findings related to the correlation and autocorrelation structures are provided.

Chapter 3 focuses on exploring the possibilities of the *BMMPP* for the modeling of real phenomena involving point processes with group arrivals. The first result in this sense is the characterization of the $BMMPP_2(K)$ by a set of moments related to the inter-event time and batch size distributions. This characterization leads to a sequential fitting approach via a moments matching method. The performance of the novel fitting approach is illustrated on both simulated and a real teletraffic data set, and compared to that of the EM algorithm. In addition, as an extension of the inference approach, the queue length distributions at departures in the queueing system $BMMPP/M/1$ is also estimated.

Unlike Chapters 2 and 3, which are devoted to the Batch Markov Modulated Poisson Process, Chapter 4 presents an extension to the two-dimensional case of the Markov modulated Poisson process (*MMPP*), motivated by real failure data in a two-dimensional context. The one-dimensional *MMPP* has been proposed for the modeling of dependent and non-exponential inter-event times (in contexts as queuing, risk or reliability, among others). The novel two-dimensional *MMPP* allows for dependence among the two sequences of inter-event times, while at the same time preserves the *MMPP* properties marginally. Such generalization is based on the Marshall-Olkin exponential distribution. Inference is undertaken for the new process through a method combining a matching moments approach and an ABC algorithm. The performance of the method is shown on simulated and real datasets representing failures of a public transport company.

To conclude, Chapter 5 summarizes the most significant contributions of this dissertation, and also gives a short description of possible research lines.

Resumen

Esta tesis está principalmente motivada por el problema de la modelización estadística mediante un tipo específico de procesos puntuales: el (Batch) Markov Modulated Poisson proces. Los procesos puntuales aparecen en una amplia variedad de situaciones en la física, la biología, la ingeniería o la economía. En general, la ocurrencia de eventos se define en relación con el contexto, pero en muchas áreas están vinculados a la ocurrencia de un evento en un momento específico. En la mayoría de las ocasiones con el objetivo de simplificar los modelos y obtener expresiones cerradas para las cantidades de interés, se asume que los tiempos entre eventos son independientes y exponencialmente distribuidos. Sin embargo, las suposiciones de independencia y exponencialidad pueden ser poco realistas y restrictivas en la práctica, y por consiguiente, se necesitan modelos más alineados con la realidad para ajustar los datos.

Los Batch Markov Modulated Poisson Processes (*BMMPP*s) son una subclase de los Batch Markovian Arrival processes (*BMAP*s), los cuales han sido ampliamente usados para la modelización de eventos correlados (tales como llegadas, fallos o eventos de riesgo). Tanto los *BMMPP*s como los *BMAP*s permiten la dependencia y no exponencialidad de los tiempos entre eventos así como las llegadas en grupo correladas, pero al mismo tiempo heredan la tratabilidad de los procesos de Poisson. Por lo tanto, resultan ser modelos adecuados para ajustar los datos con características estadísticas que difieren de los supuestos clásicos de los procesos de Poisson. A pesar de la gran cantidad de trabajos considerando las subclases de procesos *BMAP*, persisten varios problemas abiertos que son de interés y que serán considerados en esta tesis, la cual se organiza de la siguiente manera.

En el Capítulo 1, se presenta un breve recorrido teórico y metodológico que introduce los conceptos y propiedades más importantes que se necesitan para desarrollar las principales aportaciones de la tesis. Además se introducen los procesos puntuales Markovianos y sus principales propiedades.

En el Capítulo 2 se prueba la identifiabilidad del $BMMPP_m(K)$ estacionario, donde K es el tamaño máximo del grupo de llegada y m es el número de estados de la cadena de Markov subyacente. Este es un resultado deseable de cara a la inferencia. Además se prueban algunas propiedades nuevas relacionadas con las estructuras de correlación y autocorrelación.

El Capítulo 3 se centra en la exploración de las posibilidades del *BMMPP* para la modelización de fenómenos reales que involucran procesos puntuales con llegadas en grupo. El primer resultado en este sentido es la caracterización del $BMMPP_2(K)$ por un conjunto de momentos relacionados con los tiempos entre llegadas y la distribución del tamaño de los grupos. Esta caracterización lleva a un enfoque de ajuste secuencial a través de un método de ajuste por momentos. El grado de ajuste y la potencia de este nuevo enfoque se ilustran tanto con datos simulados como con una base de datos reales de teletráfico y se compara con el algoritmo EM. Además, como una extensión del enfoque de inferencia, se estiman las distribuciones de la longitud de la cola en el momento en que terminan los servicios en un sistema de cola $BMMPP/M/1$.

A diferencia de los Capítulos 2 y 3, que se dedican al Batch Markov Modulated Poisson Process, el Capítulo 4 presenta una extension al caso bidimensional del Markov Modulated Poisson process (*MMPP*), motivado por datos reales de fallos en un contexto bidimensional. El modelo unidimensional *MMPP* se ha propuesto para la modelización de tiempos entre eventos dependientes y no exponenciales (en contextos como la teoría de colas, el riesgo o la fiabilidad, entre otros). El nuevo *MMPP* bidimensional permite la dependencia entre las dos secuencias de *tiempos* entre eventos, mientras conserva las propiedades del *MMPP* marginalmente. Esta generalización se basa en la distribución exponencial bivariante Marshall-Olkin. Además se lleva a cabo la inferencia para el nuevo proceso a través de un método que combina el método de los momentos con el algoritmo ABC. Los resultados del método propuesto se muestran tanto para datos simulados como para una base de datos reales que representa los fallos de una compañía de transporte público de trenes.

Para concluir, el Capítulo 5 resume las contribuciones más significativas de esta tesis, y contiene además una breve descripción de posibles líneas de investigación.

Contents

List of Figures

List of Tables

Part I

Background on (Batch) Markov modulated Poisson processes

CHAPTER 1

Introduction

The main motivation for the dissertation is the exploration of the problem of statistical modeling using the Batch Markov modulated Poisson process (*BMMPP*). *BMMPP*s constitute a subclass of Batch Markovian arrival processes (*BMAP*), a general class of point processes defined by Neuts (1979) for the first time. This chapter presents an extensive overview of *BMMPP*s. In particular, Section 1.1 reviews basic definitions and properties in regards point processes, Markov processes and Markov renewal processes. Section 1.2 describes the Markovian arrival process (*MAP*) and its special case, the Markov modulated Poisson process (*MMPP*) first, to later extend such processes to their batch counterparts: namely, the batch Markovian arrival process (*BMAP*) and batch Markov modulated Poisson process (*BMMPP*). One of the key properties of *BMAP*s is the issue of identifiability, which has a direct effect on statistical estimation for the processes. This shall be considered in the first part of Section 1.3, which will be later devoted to a review of statistical approaches for fitting *BMAP* processes.

1.1 Point Processes

1.1.1 Real examples and motivation

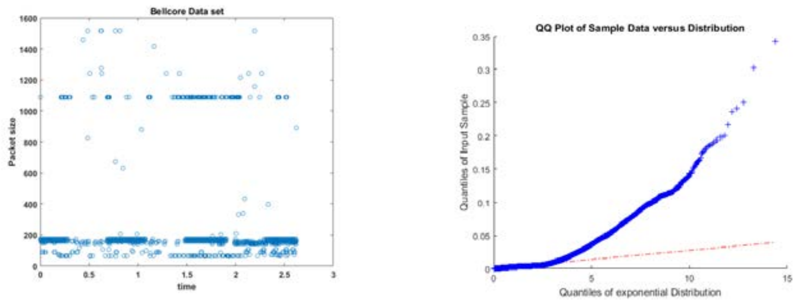
Stochastic point processes constitute a class of discrete stochastic processes, whose importance relies on their suitability for modeling a wide variety of phenomena in physics, biology, engineering, or economics, among others.

Point processes allow to model a random distribution of points in a space, which can be quite general, although in most of occasions each point represents the time and/or location of an event. As examples, consider the breackdown times of certain part of a car, the position of proteins on a cell membrane, the positions and times of earthquakes, the locations of diseased of some animal species in a given region, the instants of arrival of customers in a queue, the instants of withdrawal of items from a store, the instants of failure of a component in some system, the arrival of a packet of bytes to a computer, etc. In general, the occurrence of event is defined depending on the context.

Probabilistic models for these phenomena have been developed. Sometimes, in order to simplify the models and obtain closed form expressions for the quantities of interest, the exponentiality and/or independence of the inter-event times is assumed. However, very often these assumptions are too restrictive to be real. For example, in teletraffic contexts consider the widely used Bellcore LAN database, which is publicly available at

ftp://ita.ee.lbl.gov/html/contrib/BC.html

The database consists of four traces, each one contains a million packet arrivals seen on an Ethernet at the Bellcore Morristown Research and Engineering facility. We will focus on BC-pAug89, which began at 11:25 on August 29, 1989, and ran for about 3142.82 seconds (until one million of packets had been captured). The data file consists of two columns in ASCII format, where the first column gives the time in seconds of the packet arrival since the start of the trace, and the second column gives the Ethernet data length in bytes where all packets have at least a minimum size of 64 bytes and at most the maximum size of 1518 bytes. Figure 1.1a shows the arrival times for the first 1000 packets and the size in bytes of each one. On the other hand, Figure 1.1b shows the QQ-plot of the packet interarrival times in comparison with that of an exponential distribution. From which it can be concluded that the exponential distribution would not provide a good fit for the data.



(a) Plot of the first 1000 arrival times versus the packet sizes in bytes)

(b) Inter-arrival times QQ plot versus the QQ plot fitted by an exponential distribution

Figure 1.1: Analysis of the Bellcore dataset

Another example can be found in an insurance context, where typically the Poisson process is considered for modeling. The first data set under consideration in this context is the *Norwegian fire insurance data*, studied in Beirlant et al. (1996). This dataset represents the amounts (in Krone) and times of 9181 claims to a fire insurance company in Norway for the period 1971-1992. The second one is the *Secura Belgian Re data*, analyzed in Beirlant et al. (2006). This dataset contains 371 automobile claims made to several European insurance companies from 1988-2001. Figures 1.2 and 1.3 show the QQ plot and empirical autocorrelation function for these two data sets. In Figure 1.2 can be seen that the data has a non-negligible correlation, and therefore any risk model assuming independence between the inter-event times will not be reasonable and will produce unreliable estimates and predictions. On the other hand, from Figure 1.3 is clear that exponential model does not capture properly the quantiles structure of the dataset.

Something similar occurs in finance problems related to the modeling of the operational risk. Operational risk is in conjunction with market and credit risk what is called financial risk, which is defined as the adverse impact on performance due to different sources of uncertainty. To evaluate the operational risk, both the frequency (or the probability of a risk event occurs); as well as the severity, or the impact of risk events on the company results,

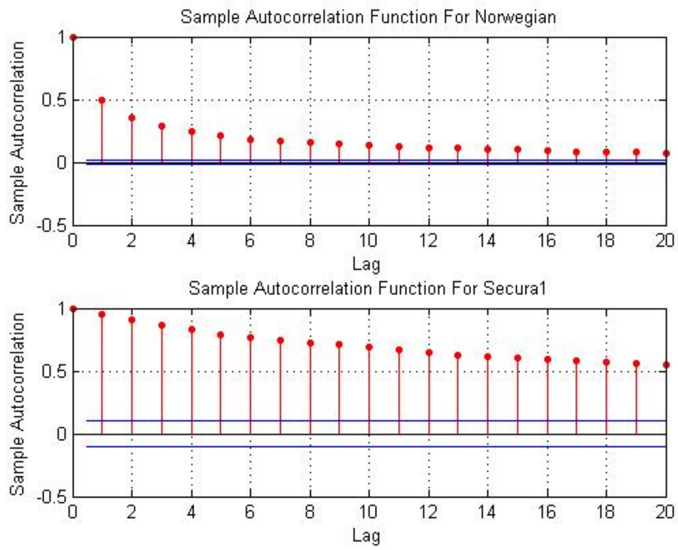


Figure 1.2: The ACF for the claim amounts in the real databases

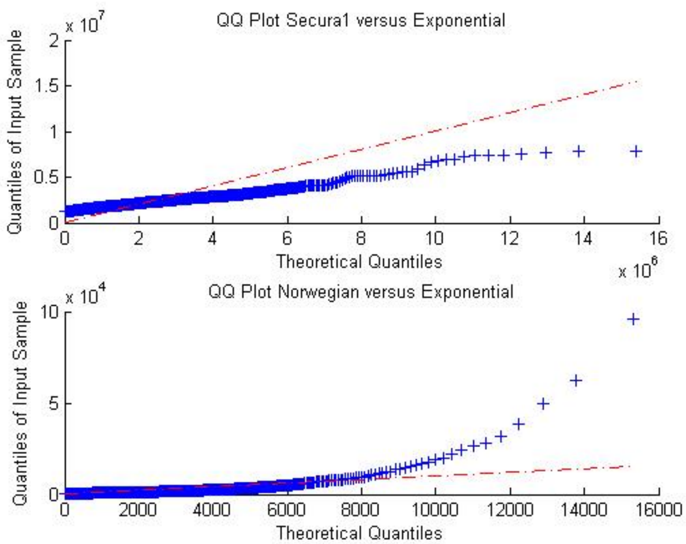


Figure 1.3: The empirical quantiles of the inter-arrival times versus those of an exponential distribution

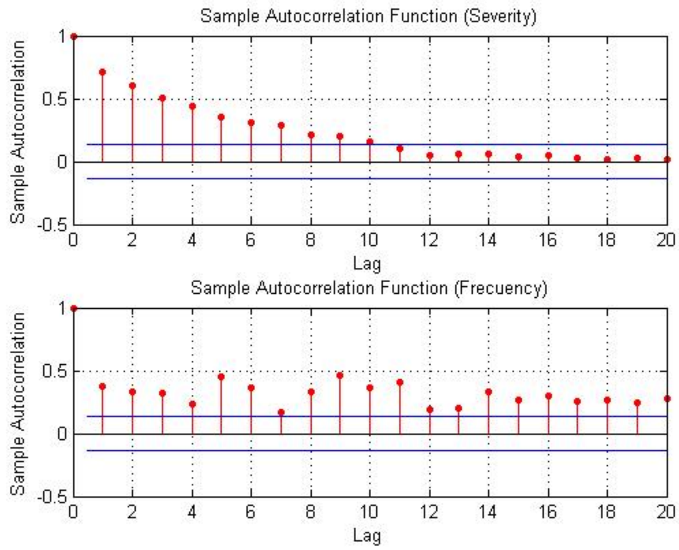
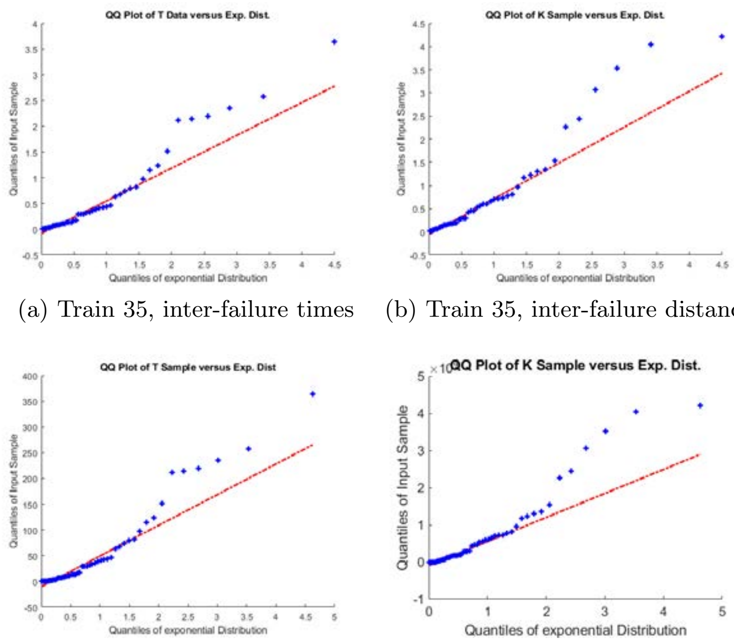


Figure 1.4: Empirical autocorrelation functions of both variables

are usually considered. This hypothesis of independence facilitates greatly the calculations, but in certain contexts it can be unrealistic. In order to illustrate it, Figure 1.4 shows the autocorrelation functions of severities and frequencies from some minority banking financial institution. The observations were taken from 30/12/93 to 29/06/2007, the trace consists of 225 observations and it is considered that there is a single type of risk event. In Figure 1.4 it can be seen that the correlation is non-negligible. Therefore, assuming that the variables are independent and identically distributed would not be appropriate for this data set.

Finally, consider an example from a reliability context, which consists of the records of the failures of two trains for about 8 years of an European transportation company. This data set was studied in Pievatolo and Ruggeri (2010) and Pievatolo et al. (2003). Figure 1.5 shows the non-exponentiality of the inter-failures times and the distance travelled between such failures distances for those two trains.

All previous datasets are examples of observed point proceses, for which the exponential distribution and the assumption of independence are not



(a) Train 35, inter-failure times (b) Train 35, inter-failure distances

(c) Train 36, inter-failure times (d) Train 36, inter-failure distances

Figure 1.5: QQ-plots of inter-failure times/distances: empirical versus exponential distribution.

valid. Next section formally defines a point process.

1.1.2 Formal definition and examples

A point process is a random distribution of points on a complete separable metric space, taking values in the non-negative integers. This metric space can be $\mathbb{R}$ if the events we want to model are the epochs at which a certain event occurs.

Let us define $\{S_n, n \geq 1\}$ as an increasing sequence of random variables that represent the time of the n^{th} event occurrence and T_n as the elapsed time between S_{n-1} and S_n. Therefore, S_n can be written as $S_n = \sum_{i=1}^{n} T_i$. This kind of time-dependent processes can also be specified by the counting process $\{N(t), t \geq 0\}$, where $N(t) = \max\{n : S_n \leq t\}$, in other words, $N(t)$ is a random variable that count the number of events occurred in the interval $(0, t]$. Note that $\{S_n \leq t\}$ if and only if $\{N(t) \geq n\}$. A counting process $\{N(t), t \geq 0\}$ is defined to possess stationary increments if for every set of time instants $t_0 = 0 < t_1 < t_2 < \cdots < t_n$, the increments $N(t_1) - N(t_0), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ are identically distributed. We will now proceed to study two important examples of point processes: one with stationary increments (the Poisson process) and another that generally does not satisfy this property (the renewal process).

Poisson process

The Poisson process is usually defined as follows.

Definition 1. *A counting process $\{N(t), t \in [0, \infty)\}$ is a homogeneous Poisson process with rates $\lambda > 0$, if all the following conditions hold:*

(1) N(0)=0.

(2) N(t) has stationary and independent increments.

(3) $P(N(h) = 1) = \lambda h + o(h)$.

(4) $P(N(h) \geq 2) = o(h)$.

where $f(h) = o(h)$ means $\lim_{h \to 0} \frac{f(h)}{h} = 0$.

A characteristic property of the Poisson process is that the inter-event times are exponentially distributed with mean $\frac{1}{\lambda}$. This property provides an alternative definition of the Poisson process and a convenient way of simulating it.

This second characterization of homogeneous Poisson process implies that, starting from an arbitrary point, the time until the r^{th} point has a gamma distribution, in other words

$$P(N(t) - N(s) = k) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^k}{k!}, \qquad 0 \leq s \leq t, \quad k \in \mathbb{Z}^+.$$

A natural and useful generalization of the Poisson process is to consider that λ depends on time. Under this assumption, the non-homogeneous

Poisson process (NHPP) is obtained. However, the NHPP assumes that inter-arrival times are also independent. Therefore, it does not seem to be a viable option in contexts such as those presented in Section 1.1.1.

Renewal processes

Another important class of point processes are the renewal processes which are obtained by assuming that the time intervals are independent of each other, but are not governed by an exponential distribution. In other words, the renewal process are point processes for which the sequence of inter-event times $(T_1, T_2, \dots)$ are independent and identically distributed random variables with an arbitrary distribution. In general it is assumed that T_1 follow a general law G_1, different from the general law G governing the rest of the inter-event times (T_i, for $i > 1$). These models are very versatile and used in practice due to their mathematical treatability that allows to obtain explicit quantities related to the performance of the associated processes. An example of this are the following studies using the renovation processes applied to reliability in manufacturing processes (Ali and Pievatolo, 2016), earthquake occurrences (Jatiningsih et al., 2019) and actuarial applications (Lveill and Hamel, 2019).

A particular case of a tractable and widely applied renewal processes is the Phase-type renewal processes which is examined below.

The PH renewal processes are renewal processes whose sequence of inter-event times are independent and follow a phase-type distribution. They have been widely used in practice in many areas, since they are analytically and algorithmically tractable models. For example, in reliability (Montoro-Cazorla and Pérez-Ocón, 2006), in queueing theory (Kim et al., 2010), and in healthcare (Marshall et al., 2015). Intuitively, the phase-type distribution can be represented as the time until absorption in a Markov process with one absorbing state. From this, it can be inferred that a PH-distribution can be written as a mixture of exponential distributions. Each of these exponential distributions represents a phase or state of the process and may or may not have the same parameter. Consider a process with m phases starting in phase i (for $i = 1, ..., m$), after an exponential sojourn time in that phase the processes jumps to phase j (for $j = 1, ..., m$). The process goes from one phase to another until it arrives to the absorbing state. In He (2014) the following algebraic definition of the PH-distribution can be found.

Definition 2. *Let X be a non-negative random variable. X is said to be PH-distributed, represented as $(\boldsymbol{\alpha}, \boldsymbol{T})$, if its cumulative distribution function is given as*

$$F_X = P(X \leq s) = 1 - \boldsymbol{\alpha} e^{\boldsymbol{T}s}\boldsymbol{e} = 1 - \boldsymbol{\alpha}\left(\sum_{n=0}^{\infty}\frac{s^n}{n!}\boldsymbol{T}^n\right)\boldsymbol{e}, \quad \textit{for } s \geq 0,$$

where $\boldsymbol{e}$ is a column vector of ones, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_m)$ is a row vector of order $m > 0$ (number of phases), where $\alpha_i \geq 0$ and $\boldsymbol{\alpha e} = 1$ and $\boldsymbol{T}$ is an invertible $m \times m$ matrix satisfying that all row sums are non-positive, $(T)_{ii} < 0$ and $(T)_{ij} \geq 0$, for $i, j \in \{1, 2, ...m\}$ and $i \neq j$.

The versatility of PH distributions, containing as particular cases the exponential and Erlang distributions, along with the possibility of approximating any positive-valued distribution for a sequence of PH-distributions (see Asmussen (2008) for a proof), justify its extensive use. However, as with the Poisson process, the assumption of independent inter-event times in the renewal processes is restrictive in practice as shown in Section 1.1.1. In Section 1.2 this assumption is relaxed by means of the Markov process.

1.1.3 Markov chains and Markov processes

A Markov chain is a collection of random variables having the property that, given the present, the future is conditionally independent of the past. A formal definition is as follow:

Definition 3. *Let $X(t), t \in T$ a stochastic process, whose values are elements of a state space S and T denotes time. Then $X(t)$ is said to be a*

Markov process if the Markov property is satisfied:

$$P\left(X(t_{k+1}) = x_{k+1} | X(t_k) = x_k, ..., X(t_0) = x_0\right) = P\left(X(t_{k+1}) = x_{k+1} | X(t_k) = x_k\right),$$

for any $0 \leq t_0 \leq ... \leq t_k \leq t_{k+1}$ *and* $x_i \in S$.

We can say that in a Markov process the future depends on the past only through the present. In other words, if the state of the stochastic process at a specific time is known, then it is possible to predict the future stochastic behavior. For this reason the Markov processes are a very suitable tools in stochastic modeling problems.

Discrete time Markov Process

Let $\{X_n, n = 0, 1, 2, ...\}$ be a discrete-time stochastic process, where X_n represents the state in which the process is at the instant of time n. Then, $\{X_n, n \geq 0\}$ is a discrete Markov process or Markov chain, if

$$P\left(X_{n+1} = j | X_n = i, X_k = x_k, 0 \leq k \leq n-1\right) = P\left(X_{n+1} = j | X_n = i\right) = p_{ij}.$$

The previous property states that knowing the present state $\{X_n\}$, the future state $\{X_{n+1}\}$ does not depend on the previous states $\{X_0, ..., X_{n-1}\}$. On the other hand, the state transition probability can be represented in a transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0j} & \cdots \\ p_{10} & p_{11} & \cdots & p_{1j} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ p_{i0} & p_{i1} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}.$$

Since the elements of the matrix $\boldsymbol{P}$ are state transition probabilities, they are non-negative. In other words for any row i, $\sum_{j=0}^{\infty} p_{ij} = 1$.

Let $p_{ij}^{(n)}$ be the probability that the process goes from state i to j after n transitions, then

$$p_{ij}^{(n)} = P\left(X_{m+n} = j | X_m = i\right), \quad \text{for } n > 0$$

and

$$p_{ij}^{(0)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

The transition probabilities satisfy the Chapman-Kolmogorov equations

$$p_{ij}^{(n)} = \sum_{k=0}^{\infty} p_{ik}^{(r)} p_{kj}^{(n-r)}, \quad 0 \leq r \leq n.$$

These equations also have a matrix version

$$\boldsymbol{P}^{(n)} = \{p_{ij}^{(n)}\} = \boldsymbol{P}^{(1)} \boldsymbol{P}^{(n-1)} = \boldsymbol{P} \boldsymbol{P}^{(n-1)} = ... = \boldsymbol{P}^n.$$

Therefore, a Markov chain is fully characterized by the transition probability matrix $\boldsymbol{P}$ and an initial probability vector $\boldsymbol{\alpha}$, where $\alpha_i = P(X_0 = i)$. We can compute the probability that the process is in state j after the n-th transition $\alpha_j^{(n)} = P(X_n = j)$ as

$$\alpha_j^{(n)} = \sum_{i=0}^{\infty} P[X_n = j, X_0 = i] P[X_0 = i] = \sum_{i=0}^{\infty} p_{ij}^{(n)} \alpha_i.$$

A classification of Markov chains is given next. Let T be the time of the first visit to state i. Then such state i can be classified as:

1. *Recurrent* if $P(T < \infty | X_0 = i) = 1$. In addition, state i is
   (a) *Null Recurrent* if $E(T|X_0 = i) = \infty$
   (b) *Positive Recurrent* if $E(T|X_0 = i) < \infty$
2. *Transient* if $P(T = \infty | X_0 = i) > 0$

Moreover, let d denote the largest integer for which $P(T = nd|X_0 = i) = 1$. If $d = 1$, then the recurrent state i is said to be aperiodic; otherwise, if $d > 1$

is called periodic with period d. On the other hand, a Markov chain is said to be irreducible if and only if all states can be reached from each other. If a Markov chain is irreducible, it can be shown that all states are either positive recurrent or transient. Finally, the stationary probability vector is defined in the following result.

Proposition 1. *For a finite, irreducible and aperiodic Markov chain, there exists a unique stationary probability vector that satisfies*

$$\lim_{n\to\infty} \boldsymbol{P}^{(n)} = \boldsymbol{e}\boldsymbol{\phi}, \quad \boldsymbol{\phi} = \boldsymbol{\phi}\boldsymbol{P} \quad \text{and} \quad \boldsymbol{\phi}\boldsymbol{e} = 1,$$

where $\boldsymbol{e}$ *is a vector of ones.*

Next section considers continuous-time Markov Processes for which transitions can occur at any $t > 0$.

Continuous-time Markov Process

A continuous-time stochastic process $\{X(t), t \in \mathrm{R}^+\}$ is a continuous-time Markov process if

$$\begin{aligned} P\left(X(t+s) = j | X(s) = i, X(u) = u, 0 \leq u \leq s\right) &= P\left(X(t+s) = j | X(s) = i\right) \\ &= p_{ij}(t). \end{aligned}$$

In other words, the future $X(s+t)$ depends on the past $X(u)$ only through the present $X(s)$.

As for the discrete process in the continuum process, the probability of passing from one state to another in a given time can be defined. Let $p_{ij}(t)$ be the probability that the Markov process will be in state j, given that it was in i, t time units ago. This quantity is called transition probability function and satisfies the Chapman-Kolmogorov equation given by

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s).$$

Let $\boldsymbol{P}(t) = [p_{ij}(t)]$ be the matrix with transition probabilities. Then, $\boldsymbol{P}(t)$ is called the transition probability matrix at time t and satisfies that all the elements of each row sum one, $\sum_j p_{ij}(t) = 1$. Using this matrix representation the Chapman-Kolmogorov equation can be rewritten as

$$\boldsymbol{P}(t+s) = \boldsymbol{P}(t)\boldsymbol{P}(s).$$

Let H_i be the time that the process spends in state i before leaving it. This time is called sojourn time and due to the Markovian property it is memoryless, or, in other words

$$P[H_i > s+t|H_i > s] = P[H_i > t], \quad s,t \geq 0.$$

Therefore, H_i is exponentially distributed with mean $1/\lambda_i$, where λ_i represents the rate at which the process leaves state i, while $\lambda_i p_{ij}$ denotes the rate of transition from state i to j. Using these two rates it is possible to define a matrix that plays the same role as the transition matrix $\boldsymbol{P}$ of the discrete-time Markov chains. This matrix, also known as infinitesimal generator, is defined as $\boldsymbol{Q} = \{q_{ij}\}_{i,j\in S}$, where

$$q_{ij} = \begin{cases} \lambda_i & \text{if } i = j \\ -\lambda_i p_{ij} & \text{if } i \neq j \end{cases}$$

and it satisfies

$$\lim_{t\to 0} \frac{\boldsymbol{P}(t) - \boldsymbol{I}}{t} = \boldsymbol{Q} \quad \text{and} \quad \frac{\boldsymbol{P}(t)}{dt} = \boldsymbol{Q}\boldsymbol{P}(t) = \boldsymbol{P}(t)\boldsymbol{Q}, \text{ for } t \geq 0,$$

where $\boldsymbol{P}(t)$ is the probability transition matrix of the process. From these results when state-space is finite, it can be deduced that

$$\boldsymbol{P}(t) = \boldsymbol{Q}t.$$

If we consider the continuous-time Markov process $\{X(t), t \geq 0\}$ only at the instants upon which a state transition occurs $(t_0, t_1, t_2, ...)$, then, it is possible to construct a Markov chain $\{X_n, n \geq 0\}$. The values of this Markov chain will be the state of $\{X(t)\}$ immediately after the transition at time t_n. Using this construction the states of a Markov process can be classified by the classification provided by the embedded Markov chain.

Definition 4. *Let* $\{X(t), t \geq 0\}$ *be a continuous-time Markov process and* $\{X_n, n \geq 0\}$ *its associated embedded discrete-time Markov chain. Then*

1. $\{X(t)\}_{t>0}$ *is irreducible if and only if* $\{X_n\}_{n\geq 0}$ *is irreducible.*
2. *A state* i *is recurrent/transient for* $\{X(t)\}_{t>0}$ *if and only if it is recurrent/transient for* $\{X_n\}_{n\geq 0}$.

For irreducible and positive recurrent Markov process is possible to obtain the follow limiting probability

Proposition 2. *Let* $\{X(t), t \geq 0\}$ *be an irreducible and positive recurrent Markov process, then there exists an unique stationary probability vector* $\boldsymbol{\pi}$ *that is independent of the initial state and satisfies*

$$\pi_j = \lim_{t\to\infty} P[X(t) = j], \quad \textit{and} \quad \boldsymbol{\pi}\boldsymbol{Q} = 0.$$

1.1.4 Markovian renewal processes

Markov renewal processes are stochastic processes in which the transitions from one state to another one occur according to a Markov chain, and the time between two successive state transitions is a random variable with a general distribution that depends on the current state as well as the successive transition state. A stochastic process that combines renewal processes and Markov chains is called a Markov renewal process. A formal definition is given next. Let $\{X_n, n \geq 0\}$ be a stochastic process on state space S, where X_n records the state of the process at n-th event occurrence and S_n is a random variable that denotes the time at which that event occurs.

Definition 5. *Let* $\{X_n, S_n, \}_{n\geq 0}$ *and* $K_{i,j}(t)$ *be a bivariate stochastic process and the probability that the process passes from state* i *to state* j *in a period of time* t *respectively. The process is called a Markov renewal process with state space* S *if*

$$K_{i,j}(t) \quad = \quad P[X_{n+1} = j, S_{n+1} - S_n \leq t | X_0, ..., X_n = i; S_0, ..., S_n]$$

$$= \quad P[X_{n+1} = j, S_{n+1} - S_n \leq t | X_n = i].$$

In other words, if we write this values in a matrix way as $\boldsymbol{K}(t) = [K_{i,j}(t)]$, then $\boldsymbol{K}(t)$ is known as the semi-Markov kernel of the Markov renewal process $\{X_n, S_n\}$. If P^*_{ij} is defined as

$$P^*_{ij} = \lim_{t \to \infty} K_{ij}(t),$$

then P^* is the transition probability matrix of the Markov chain $\{X_n, n \geq 0\}$ with state space S. The Markov chain $\{X_n, n \geq 0\}$ governs transitions between states, therefore it is said to be an embedded Markov chain. For more details on Markov Renewal processes, we refer the reader to Nakagawa (2011).

Next section introduces the stochastic process we are dealing with in this dissertation. It is a specific class of Markov renewal process for which the inter-arrival times are non-exponential and dependent. Such a process is named Batch Markovian arrival process, a class of processes that generalize the renewal processes with phase-type distributions and allows for dependence among inter-event times via its markovian structure.

1.2 Fundamentals on Batch Markovian arrival process

The Batch Markovian Arrival Process (*BMAP*), as stated above, were introduced by Neuts (1979) as a versatile Markovian point process. However, its matrix and more tractable notation is due to Lucantoni (1991). *BMAP*s have the important characteristic of being dense in the family of all stationary point processes. Asmussen and Koole (1993) proved that it is possible to find a sequence of *BMAP*s that converges to a given point process. On the other hand, *BMMPP* is simpler a subclass with fewer parameters than *BMAP* and for some situations a *BMMPP* is enough to capture the complexity of the problem. Therefore, both of them have been widely considered in a number of real-life contexts where dependent arrivals are commonly observed. For example, in reliability (Montoro-Cazorla and Pérez-Ocón, 2006; Montoro-Cazorla and Pérez-Ocón, 2014b,a; Rodríguez et al., 2016b; Rodríguez et al., 2015), teletraffic (Kang et al., 2002; Casale et al., 2010; Wang et al., 2015), insurance (Landriault and Shi, 2015; Li and Ren, 2013), and weather forecasting (Ramírez-Cobo et al., 2014b). On the other hand, the versatility and tractability of *BMAP* and *BMMPP* make them suitable tools for modeling the bursty arrival processes commonly arising in computer and communications applications, specially in queuing models, because it provides a way to model more complex arrival systems.

BMMPP is the batch counterpart of the widely known Markov modulated Poisson process (*MMPP*), which is a sub class of the Markov arrival process (*MAP*). Therefore, continuing with the structure followed in the description of the previous processes, the *MAP* and the *MMPP* will be introduced first, to then, move on to the description of *BMAP* and *BMMPP* which are more general processes.

1.2.1 The Markovian arrival process and the Markov modulated Poisson process

The *MAP* was introduced by Neuts (1979) as a versatile Markovian point process. However, its matrix and more tractable notation is due to Lucantoni (1991). Both *MAP* and *MMPP* can be seen as a matrix generalization of the Poisson process inheriting its tractability and extending its capabilities.

Let consider a Poisson process with rate λ and denotes as $N(t)$ the number of arrivals in $(0, t]$. Then $\{N(t)\}_{t\geq 0}$ is a Markov process on the state space $S = \mathrm{Z}^+$ with infinitesimal generator

$$\boldsymbol{G} = \begin{pmatrix} d_0 & d_1 & 0 & 0 & \cdots \\ 0 & d_0 & d_1 & 0 & \cdots \\ 0 & 0 & d_0 & d_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where, $d_0 = -\lambda$ and $d_1 = \lambda$.

As mentioned above, it is desirable to find a process that allows for non-exponential times between event occurrences, dependence between inter-event times but still preserving an underlying Markovian structure. A process that fulfills these specifications is the m-states MAP, denoted as MAP_m, which is constructed by generalizing the above Poisson process. Let $\{J(t), N(t)\}$ be a two-dimensional Markov process on the state space $S = \mathrm{Z}^+ \times [0, m]$; where $J(t)$ is an irreducible and continuous Markov process with state space $S = \{1, ..., m\}$ and $N(t)$ is a counting process that quantifies the number of events occurring in the interval $(0, t]$. While the infinitesimal generator of the process is given by

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{D_0} & \boldsymbol{D_1} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \boldsymbol{D_0} & \boldsymbol{D_1} & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D_0} & \boldsymbol{D_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where both $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ are $m \times m$ square matrix, being nonnegative all the off-diagonal elements of $\boldsymbol{D_0}$ and all the elements of $\boldsymbol{D_1}$.

The MAP_m behaves as follows: the initial state $i_0 \in S$ is generated according to an initial probability vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$, and at the end of an exponentially distributed sojourn time in state i, with mean $1/\lambda_i$, two types of transitions can occur. On one hand, with probability p_{ij0}, no event occur and there is a transition from state i to state j (with $i \neq j$), in other words, the process changes state. On the other hand, with probability p_{ij1} an event occurs and there is a transition from i to state j (where it could be $i = j$), in other words the process can change state or stay in the same state. Therefore, the transition probabilities satisfy

$$\sum_{j=1, j\neq i}^{m} p_{ij0} + \sum_{j=1}^{m} p_{ij1} = 1, \qquad \text{for all } i \in S. \tag{1.1}$$

Hence, any MAP_m can be characterized by the initial probability vector, $\boldsymbol{\alpha}$, and the rate matrices $\boldsymbol{D_0}$, and $\boldsymbol{D_1}$, which govern the transitions without and with event occurrences, respectively, and are defined as

$$\begin{aligned} (\boldsymbol{D_0})_{ii} &= -\lambda_i, & i \in S, \\ (\boldsymbol{D_0})_{ij} &= \lambda_i p_{ij0}, & i,j \in S, i \neq j, \\ (\boldsymbol{D_1})_{ij} &= \lambda_i p_{ij1}, & i,j \in S. \end{aligned} \tag{1.2}$$

On the other hand, in the stationary processes, which are the focus of this thesis, the initial probability vector, $\boldsymbol{\alpha}$, coincides with the stationary probability vector, $\boldsymbol{\pi}$, of the Markov process with generator $\boldsymbol{Q} = \boldsymbol{D_0} + \boldsymbol{D_1}$ (see Chakravarthy (2001)). Consequently $\boldsymbol{\pi} = (\pi_1, ..., \pi_m)$, is the unique (positive) probability vector satisfying $\boldsymbol{\pi Q} = \boldsymbol{0}$ and $\boldsymbol{\pi e} = 1$, where $\boldsymbol{0}$ and $\boldsymbol{e}$ are columns vector of zeros and ones respectively. Hence, π_i is the stationary probability that the process is in state i.

The MAP_m can be viewed as a Markov renewal process. If we denote T_n as the time between the $(n-1)$-th and the n-th occurrences and S_n the state of the process when n-th event occurs, then, $\{S_n, T_n\}_{n=1}^{\infty}$ is a Markov renewal process and $\{S_n\}_{n=1}^{\infty}$ is a Markov chain whose transition probability matrix is

$$\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}. \tag{1.3}$$

On the other hand, its stationary probability vector $\boldsymbol{\phi}$ is computed solving the equation $\boldsymbol{\phi} = \boldsymbol{\phi P^*}$, whose solution is given by

$$\boldsymbol{\phi} = (\boldsymbol{\pi D_1 e})^{-1}\boldsymbol{\pi D_1}, \tag{1.4}$$

see Ramírez-Cobo et al. (2010) for a proof.

In practice, it is usually not possible to observe the sequence of states and only a sequence of inter-event times $\boldsymbol{t} = (t_1, t_2, ..., t_n)$ is observed. Hence, it is important to study the properties of inter-event times. In the stationary case the sequence of random variables $\{T_n\}_{n\geq 1}$ are identically distributed. Let T denote time between two successive events, then T follows a phase-type distribution with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}\}$ (see Latouche and Ramaswami (1999)). Therefore, the cumulative distribution function of T is

$$F_T(t) = 1 - \boldsymbol{\phi} e^{\boldsymbol{D_0} t} \boldsymbol{e} \tag{1.5}$$

and its moments are

$$\mu_n = E[T^n] = n! \boldsymbol{\phi}(-\boldsymbol{D_0})^{-n} \boldsymbol{e}. \tag{1.6}$$

As mentioned above, one of the most important properties of the MAP_m is that it allows for a dependence structure between the inter-event times. In particular

$$\rho_T(l) = \mathrm{Corr}(T_1, T_{l+1}) = \frac{\boldsymbol{\pi}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}]^l(-\boldsymbol{D_0})^{-1}\boldsymbol{e} - \mu_1}{2\boldsymbol{\pi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e} - \mu_1}. \tag{1.7}$$

Finally, Rodríguez et al. (2016a) give the Laplace-Stieltjes transform of the n first consecutive inter-event times of a non-stationary MAP_m. Taking $\boldsymbol{\pi}$ as the initial probability vector, the Laplace-Stieltjes transform for the stationary MAP_m is given by

$$f_T^*(s_1, ..., s_n) = \boldsymbol{\phi}(s_1\boldsymbol{I} - \boldsymbol{D_0})^{-1}\boldsymbol{D_1}...(s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1}\boldsymbol{D_1}\boldsymbol{e}. \tag{1.8}$$

The Markov modulated Poisson process ($MMPP_m$) is an special subclass of the MAP_m in which the state changes only take place when no events occur. Therefore, the only change in the formulation given in (1.2) is that $p_{ij1} = 0$ for $i \neq j$, which implies that $\boldsymbol{D_1}$ is a diagonal matrix. This simplification will allow $MMPP_m$ to be identifiable, a very desirable property in terms of inference, while inheriting much of the versatility that MAP_m has.

Example: Stationary $\boldsymbol{MMPP_2}$. The simplest case of $MMPP_m$ is the stationary $MMPP_2$, hence $m = 2$ and state space is $S = \{1, 2\}$. The behavior of this process is illustrated by means of the transition diagram of Figure 1.6, where all the different transitions that can occur in this process appear.

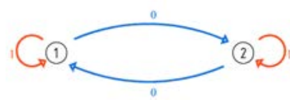


Figure 1.6: Transition diagram for the $MMPP_2$. Here 0 and 1 illustrate moves without and with arrivals, respectively.

In this particular case the rate matrices are given by

$$\boldsymbol{D_0} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{120} \\ \lambda_2 p_{210} & -\lambda_2 \end{pmatrix}, \quad \boldsymbol{D_1} = \begin{pmatrix} \lambda_1 p_{111} & 0 \\ 0 & \lambda_2 p_{221} \end{pmatrix},$$

where $p_{111} = 1 - p_{120}$ and $p_{221} = 1 - p_{210}$. In the next section the batch counterpart of MAP_m and $MMPP_m$ will be analyzed.

1.2.2 The batch Markovian arrival process and the batch Markov modulated Poisson process

In the same way that MAP_m is defined as an extension of a Poisson process, $BMAP_m(K)$, where K is the maximum batch arrival size, can also be defined as an extension of a Poisson process with batch events. Let consider a Batch Poisson process with rate λ and with p_k being the probability that the batch size is equals to k, with $k \in \mathrm{N}$. If $N(t)$ is the number of arrivals in $(0, t]$, then $\{N(t)\}_{t\geq 0}$ is a Markov process on the state space $S = \mathrm{Z}^+$ with infinitesimal generator

$$\boldsymbol{G} = \begin{pmatrix} d_0 & d_1 & d_2 & d_3 & \cdots \\ 0 & d_0 & d_1 & d_2 & \cdots \\ 0 & 0 & d_0 & d_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where, $d_0 = -\lambda$ and $d_k = p_k\lambda$. The functioning of this process is as follows: after an exponential sojourn time (with mean $1/\lambda$) in state i, the process jumps to state $i + k$ with probability p_k, corresponding the transition to an arrival with batch size k.

The $BMAP_m(K)$ is constructed by generalizing the above batch Poisson process. This construction allows for non-exponential times between the arrivals of batches, but still preserving an underlying Markovian structure. $\boldsymbol{G}$ now is the infinitesimal generator of a Markov process $\{J(t), N(t)\}$ on the state space $S = \mathrm{N} \times \{\mathrm{N} \cap [0, m]\}$ and has the following structure:

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{D_0} & \boldsymbol{D_1} & \boldsymbol{D_2} & \boldsymbol{D_3} & \cdots \\ \boldsymbol{0} & \boldsymbol{D_0} & \boldsymbol{D_1} & \boldsymbol{D_2} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D_0} & \boldsymbol{D_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $\boldsymbol{D_k}$, for $k \geq 1$, are $m \times m$ non-negative matrices and $\boldsymbol{D_0}$ is a $m \times m$ matrix with negative diagonal and non-negative off-diagonal elements. The irreducible infinitesimal generator for the $BMAP_m(K)$ is defined as

$$\boldsymbol{Q} = \sum_{k=0}^{\infty} \boldsymbol{D_k}.$$

As in the MAP_m, the stationary probability vector $\boldsymbol{\pi} = (\pi_1, ..., \pi_m)$, in the $BMAP_m(K)$ satisfies that $\boldsymbol{\pi Q} = \boldsymbol{\pi}$ and $\boldsymbol{\pi e} = 1$.

The $BMAP_m(K)$ behaves as follows: at the end of an exponentially distributed sojourn time in state i, with mean $1/\lambda_i$, two possible state transitions can occur. First, with probability p_{ij0} no arrival occurs and the process enters in a different state $j \neq i$. On the other hand, with probability p_{ijl}, with $1 \leq l \leq K$, there will be a transition to state j (where it could be $i = j$) with a batch arrival of size l. The transition probabilities satisfy

$$\sum_{j=1, j\neq i}^{m} p_{ij0} + \sum_{k=1}^{K} \sum_{j=1}^{m} p_{ijl} = 1, \qquad \text{for all } i \in S.$$

Like the MAP_m, the $BMAP_m(K)$ can be characterized by the rate matrices $\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K}$, where the definition of these matrices for the batch process is

$$\begin{aligned}
(\boldsymbol{D_0})_{ii} &= -\lambda_i, && i \in S, \\
(\boldsymbol{D_0})_{ij} &= \lambda_i p_{ij0}, && i, j \in S, i \neq j, \\
(\boldsymbol{D_l})_{ij} &= \lambda_i p_{ijl}, && i, j \in S, 1 \leq l \leq K.
\end{aligned} \tag{1.9}$$

If we denote by S_r the state of the $BMAP_m(K)$ at the time of the r-th, B_r the batch size of that arrive and T_r the time between the $(r-1)$-th and r-th events, then the process $\{S_{r-1}, \sum_{i=1}^{r} Ti, B_r\}_{r=1}^{\infty}$, associated to the $BMAP_m(K)$, is a Markov renewal process. Note that in the process with batch arrivals, both the times between arrivals and the states of the underlying process coincide with those of the single arrivals process. Therefore, in the $BMAP_m(K)$, $\{Sr\}_{r=0}^{\infty}$ is also a Markov chain, but its transition matrix, which for the MAP_m is given by (1.3), for the $BMAP_m(K)$ becomes

$$\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D},$$

where $\boldsymbol{D} = \sum_{l=1}^{\infty} \boldsymbol{D_l}$. Note that when $K = 1$ (the MAP_m) $\boldsymbol{D} = \boldsymbol{D}_1$.

On the other hand, T_is are also phase-type distributed with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}\}$ but for the $BMAP_m(K)$ we have to replace $\boldsymbol{D}$ by $\boldsymbol{D_1}$ in (1.4) to obtain the stationary probability vector of $\boldsymbol{P^*}$:

$$\boldsymbol{\phi} = (\boldsymbol{\pi}\boldsymbol{D}\boldsymbol{e})^{-1}\boldsymbol{\pi}\boldsymbol{D}. \tag{1.10}$$

Note that for $K = 1$ (1.4) and (1.10) coincide. Therefore the expressions obtained for the MAP_m for the cumulative distribution of T (1.5), its moments (1.6) and the auto-correlation function of the inter-event times (1.7) are also valid for the $BMAP_m(K)$.

On the other hand, Rodríguez et al. (2016c) provide the Laplace-Stieltjes transform (LST) of the n first inter-event times and batch sizes of a stationary $BMAP$ (which is the same for the $BMMPP_m(K)$):

$$f^*_{T,B}(\boldsymbol{s}, \boldsymbol{z}) = \boldsymbol{\phi}(s_1\boldsymbol{I} - \boldsymbol{D_0})^{-1}\xi(z_1)...(s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1}\xi(z_n)\boldsymbol{e}, \tag{1.11}$$

where $\boldsymbol{s} = (s_1, ..., s_n)$, $\boldsymbol{z} = (z_1, ..., z_n)$ and $\xi(z_i) = \sum_{k=1}^{K} \boldsymbol{D_k} z_i^k$, for $i = 1, ..., n$.

From the Laplace-Stieltjes transform given by Rodríguez et al. (2016c), it can be proved that the mass probability function of the stationary batch size, B, is

$$P(B = k) = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D}^k\boldsymbol{e}, \quad \text{for } k = 1, ..., K,$$

from which the moments of B can be computed as

$$\beta_r = E[B^r] = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_r^*}\boldsymbol{e}, \quad \text{for } r \geq 1, \tag{1.12}$$

where $\boldsymbol{D_r^*} = \sum_{k=1}^{K} k^r \boldsymbol{D_k}$. Another quantity of interest obtained from the LST is the autocorrelation function of the batch sequence given by

$$\rho_B(l) = \text{Corr}(B_1, B_{l+1}) = \frac{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D}]^{l-1}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e} - \beta_1^2}{\sigma_B^2}, \tag{1.13}$$

where β_1 and $\sigma_B^2 = \beta_2 - \beta_1^2$ are computed from (1.12).

A measure of interest in the $BMAP_m(K)$ is the expected number of event occurrences per unit of time, which is called the fundamental arrival rate and is given by

$$\lambda^* = \boldsymbol{\pi} \sum_{l=1}^{k} l \boldsymbol{D_l} \boldsymbol{e}.$$

Concerning the counting process $\{N(t), t \geq 0\}$, define the $m \times m$ matrix $\boldsymbol{P}(n,t) = P_{ij}(n,t)_{n \in \mathrm{N}, t \geq 0}$, whose elements are

$$P_{ij}(n,t) = P(N(t) = n, J(t) = j | N(0) = 0, J(0) = i).$$

This matrix, $\boldsymbol{P}(n,t)$, represents the conditional probability of n event occurrences in the interval $(0,t]$ and the underlying Markov process is in state j at time t, given that at initial time there was not events and the state was i. Then,

$$P(N(t) = n | N(0) = 0) = \boldsymbol{\phi}\boldsymbol{P}(n,t)\boldsymbol{e}$$

and the expected number of event occurred up to time t can be computed from the first factorial moment of the counting process:

$$M_1(t) = \sum_{n=0}^{\infty} n\boldsymbol{P}(n,t),$$

for more details, see Chakravarthy (2010) and Neuts and Li (1996).

In the same way that $MMPP_m$ was defined as a subclass of the MAP_m, $BMMPP_m$ is a special sub-class of the $BMAP_m$ in which the state changes only take place when no events occur. In other words, $\boldsymbol{D_k}$ for $1 \leq k \leq K$ are all diagonal matrices. Therefore, its formulation is expressed as

$$\begin{array}{rcll}
(\boldsymbol{D_0})_{ii} & = & -\lambda_i, & i \in S, \\
(\boldsymbol{D_0})_{ij} & = & \lambda_i p_{ij0}, & i,j \in S, i \neq j, \\
(\boldsymbol{D_l})_{ii} & = & \lambda_i p_{iil}, & i \in S, 1 \leq l \leq K, \\
(\boldsymbol{D_l})_{ij} & = & 0, & i \neq j, 1 \leq l \leq K.
\end{array} \tag{1.14}$$

In Chapter 2 will be proved that, like the $MMPP_m$ and unlike the $BMAP_m(K)$, the $BMMPP_m(K)$ is identifiable. As with the $MMPP_m$ this simplification will allow us to proof the identifiability of the $BMMPP_m$ in Chapter 2 of this dissertation.

1.3 Statistical estimation for Batch Markovian arrival process

Both *BMMPP*s as *BMAP*s have gained widespread use in stochastic modeling due to their ability to describe a wide range of situations. Hence, it is of interest to consider statistical inference for such models. A paramount aspect when studying statistical inference for stochastic process is the identifiability property which implies that the process is characterized by a unique representation. Section 1.3.1 addresses this theoretical problem for *BMAP*s and *BMMPP*s.

1.3.1 The identifiability issue for *BMAP*s and *BMMPP*s

Both (B)*MAP*s and (B)*MMPP*s are highly-parametrized models where, in practice, only inter-event times and batch sizes (in the case of the *BMAP* and *BMMPP*) are usually observed. In general, the underlying Markov chain transitions are not available, and therefore, the observed data for *BMMPP*s and *BMAP*s can be viewed as generated from a hidden Markov process, see Ephraim and Merhav (2002).

When dealing with inference for hidden Markov processes, it is common that the likelihood function does not possess a unique global maximum, or in other words, the process is non-identifiable. Several works have considered the identifiability issue for different subclasses of *BMAP*s. For example, from the results in Ito et al. (1992) and via a uniformization techinique, Rydén (1996b) studied the identifiability of the $MMPP_m$ and phase type distributions. From these works it can be concluded that while the $MMPP_m$ is identifiable (up to permutations of states), the phase-type are not. The non-identifiability of the phase-type distribution can be easily illustrated if we consider that any two-state phase-type distribution can be represented as a two-state Coxian distribution, which is not identifiable. Also, Green (1998) and Bean and Green (1999) investigated when a *MAP* is a Poisson process, while He and Zhang (2006, 2008, 2009) used the so-called spectral polynomial algorithm to construct Coxian representations for phase-type distributions whose generators have only real eigenvalues. On the other hand, Bodrog et al. (2008) provided a canonical and unique representation for the MAP_2. Concerning the MAP_2, Ramírez-Cobo et al. (2010) proved its lack of identifiability and provided the conditions for which two different MAP_2 representations are equivalent. The results in Ramírez-Cobo et al. (2010) were partially extended to the stationary MAP_3 by Ramírez-Cobo and Lillo (2012) and to the non-stationary MAP_2 by Rodríguez et al. (2016a). On the other hand, Ramírez-Cobo et al. (2014a) studied how the lack of a unique representation for the MAP_2 can affect the statistical estimation of the $MAP_2/G/1$ queueing system.

In the previous works, the identifiability of the process is defined in terms of the observable quantities (that is, the inter-event times) distribution, as follows.

Definition 6. *Let $\mathcal{B}$ be a representation of a* MAP_m *as in (1.2) and let T_n denote the time between the occurrence of the $(n-1)$-th and n-th events. Then, $\mathcal{B}$ is said to be identifiable if there does not exist a different parametrization $\widetilde{\mathcal{B}}$, such that*

$$(T_1, ..., T_n) \stackrel{d}{=} (\widetilde{T}_1, ..., \widetilde{T}_n), \quad \textit{for all } n \geq 0,$$

where $\widetilde{T}_i$ defined in analogous way as T_i, and where $\stackrel{d}{=}$ denotes equality in distribution.

Since in this dissertation we study the modeling of real phenomena by the *BMMPP*, it is of interest to investigate the identifiability of these processes first. To such aim, Definition 6 is generalized as follows,

Definition 7. *Let $\mathcal{B}$ be a representation of a* $\mathrm{BMMPP}_m(K)$ *as in (1.14) and let T_n and B_n denote the time between the $(n-1)$-th and n-th events occurrences, and the batch size of the n-th event occurrence, respectively. Then, $\mathcal{B}$ is said to be identifiable if there does not exist a different parametrization $\widetilde{\mathcal{B}}$, such that*

$$(T_1, ..., T_n, B_1, ..., B_n) \stackrel{d}{=} (\widetilde{T}_1, ..., \widetilde{T}_n, \widetilde{B}_1, ..., \widetilde{B}_n), \quad \textit{for all } n \geq 0,$$

where $\widetilde{T}_i$ and $\widetilde{B}_i$ are defined in analogous way as T_i and B_i, and where $\stackrel{d}{=}$ denotes equality in distribution.

As far as we know, the only work where the identifiability of *BMAP* or *BMMPP* has been studied is Rodríguez et al. (2016c); in particular, the non-identifiability of $BMAP_2(k)$, for $k \geq 2$ is proved. The proof, based on Ramírez-Cobo et al. (2010), consists on the construction of an equivalent $BMAP_2(k)$ to a given one, via a decomposition of the original process into k $BMAP_2(2)$s. Concerning the *BMMPP*, Chapter 2 of this dissertation focuses on the identifiability or this process. As it will be shown, the proven identifiability for the *BMMPP* turns out relevant for estimation purposes.

The remainder of this section deals with the different statistical estimation techniques that have been considered in the literature for the *BMAP*.

1.3.2 Likelihood-based approaches

This section provides an overview of the likelihood-based estimation methods used for estimating *BMAP*-related processes. The first and natural approach is to direct maximize the likelihood function for the observable sample. If $\boldsymbol{t} = (t_1, \ldots, t_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ represent the sequence of observed inter-event times and the associated batch sizes, then the likelihood function is given by

$$f_{\{\boldsymbol{D_0},\ldots,\boldsymbol{D_k}\}}(\boldsymbol{t}, \boldsymbol{b}) = \boldsymbol{\phi} e^{\boldsymbol{D_0} t_1} \boldsymbol{D_{b_1}} \ldots e^{\boldsymbol{D_0} t_n} \boldsymbol{D_{b_n}} \boldsymbol{e}. \tag{1.15}$$

Note that if we are considering a *MAP*, then $\boldsymbol{D_{b_i}} = \boldsymbol{D_1}$, for all i. Carrizosa and Ramírez-Cobo (2014) considered the direct maximization of (1.15) in the simplest and most tractable case of the MAP_2. For this, the authors report several numerical problems related to the evaluation of the function, which may turn out impossible when the sample size n grows. The reason for this, which relies in expression (1.15), is explained in detailed in the paper.

Because of the numerical trouble, the EM algorithm, see Dempster et al. (1977), which uses the *complete* likelihood function, has been widely considered instead. At each iteration of the algorithm there are two steps: the first is the expectation (E) and the other one is maximization (M). In particular, if X denotes the observable data, and Y represents the unobservable observations, (or latent variables), then, the algorithm first picks a starting value θ_0 and, for $j \geq 1$, it repeats the following steps until convergence is achieved.

(1) (E-step) Calculate

$$J(\theta|\theta^j) = E[\log f(X, Y|\theta)|Y, \theta^j]$$

(2) (M-step) Find the parameter that maximizes

$$\theta^{j+1} = arg \max_{\theta} J(\theta|\theta^{j})$$

The E-step consists in finding the distribution of the unobserved data, given the observed data set and the value of the parameters obtained in the previous step θ^j (or the initial values in the first step). While the M-step re-estimates the parameters in each step using a maximum-likelihood approach.

The EM algorithm has been proposed for inference of the *BMAP* and *BMMPP* in a number of papers, see for example Asmussen et al. (1996), Rydén (1996a), Klemm et al. (2003), Breuer (2002), Buchholz (2003), Ephraim and Roberts (2008), Okamura et al. (2009) and Horváth and Okamura (2013). It is known that the EM algorithm can be slow when the starting solution is not close to the true one. Because of this, several adaptations to the algorithm have been proposed in the context of *BMAP*s. For example, Buchholz (2003) reduced the time complexity of the algorithm by applying an uniformization technique. On the other hand, Okamura and Dohi (2009) presented a novel EM algorithm using a hyper-Erlang distribution, which does not need to compute the matrix exponential. This adaptation significantly improved the convergence of the algorithm. Also, Okamura et al. (2013) provided a deterministic annealing for the EM algorithm for a *MAP* in order to relax the local maximum convergence of algorithm. In Chapter 3 of this dissertation the EM algorithm is compared with a novel moments matching method for the $BMMPP_2(K)$.

1.3.3 Moments-based methods

The method of moments consists in solving a system of equations that equal a set of theoretical moments to their empirical counterparts. Specifically, let $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_p\}$ represent the unknown parameters characterizing our model. Then, for a sample of size n,

$$\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} t_i^j,$$

represents the $j-$th sample moment and

$$\mu_j \equiv \mu_j(\boldsymbol{\theta}) = E[T^j]$$

denotes the j-th theoretical moment. Then, the parameter estimates, $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_p\}$ is defined as the solution of the equations system given by

$$\hat{\mu}_j = \mu_j(\hat{\boldsymbol{\theta}}),$$

for $j = 1, \ldots, p$.

This simple method has turned out very convenient in the context of the *BMAP* since some sub-type of processes are completely characterized by a set of moments related to the inter-event time distribution. For some references, we refer the reader to Horváth and Telek (2002), Riska et al. (2004), Telek and Horváth (2007), Eum et al. (2007), Bodrog et al. (2008), Casale et al. (2010), Kriege and Buchholz (2011), Ramírez-Cobo et al. (2014b), Carrizosa and Ramírez-Cobo (2014) and Rodríguez et al. (2015). It is our experience that this method is fast, performs well in practice, and avoids the use of intractable likelihood functions.

In Chapter 3 of this dissertation we illustrate the convenience of the method of moments for the case of $BMMPP_2(K)$, for which the set of moments characterizing the process are first found. As it will be seen, this theoretical task (finding the moments that characterizes the distribution) is non trivial and indeed, it remains an open question for the majority of *BMAP*s.

1.3.4 Bayesian approaches

Several authors have also considered Bayesian approaches to estimate both the *MMPP* and the *MAP*. Concerning the Markov chain Monte Carlo

(MCMC) methods, Scott (1999) developed a Gibbs sampler algorithm for the $MMPP_2$. The algorithm allows to sample from the posterior distribution of the model parameters given the observed event times. Scott and Smyth (2003) estimated the $MMPP$ parameters by a rapidly mixing Markov chain Monte Carlo algorithm and apply it to a data set containing click rate data for individual computer users browsing through the World Wide Web. Fearnhead and Sherlock (2006) derived a Gibbs sampler which samples from the exact distribution of the hidden Markov chain in a $MMPP$. On the other hand, Ramírez-Cobo et al. (2017) considers an exact Gibbs sampler to sample for the posterior distribution of the model parameters. For designing the sampler, the canonical forms of the process obtained by Bodrog et al. (2008) turn out crucial.

In general, MCMC methods have proven to provide accurate approximations for $BMAP$-related processes characterized by a small number of parameters, as in the previous references. Up to our knowledge, MCMC for general $BMAP_m(K)$ have not been considered in the literature. The computational intensity inherent to Bayesian approaches with growing number of parameters, the complexity of the likelihood functions as well as the elicitation of prior distributions may be some reasons for this fact. The Approximate Bayesian Computation (ABC) is an alternative Bayesian strategy to MCMC method that avoids the evaluation of a complicated or unknown likelihood function, see for example Pritchard et al. (1999).

The Approximate Bayesian computation (ABC) is a mathematically well-founded algorithm applied in a wide variety of fields and as stated before, the method bypasses the evaluation of the likelihood function. Instead, the ABC is based on random simulations from the likelihood function at a number of iterations, where comparisons between empirical moments of the given sample and those from the generated sample are made at each iteration. Parameters leading to samples whose moments differ from those of the given sample are discarded. This idea is illustrated in Algorithm 1 (simpler version of ABC), and detailed in, for example in Pritchard et al. (1999) and Marin et al. (2012).

A description of Algorithm 1 is given next. First, a set of parameter

Algorithm 1: **ABC algorithm (simpler version)**

1. Let $\boldsymbol{t}$ represent a given sample of the process whose likelihood in intractable or unknown.
2. Fix a set of summary statistics, ν.
3. Fix a distance measure $\rho\{\cdot,\cdot\}$ and a tolerance parameter $\epsilon \geq 0$.
4. **for** $i = 1$ **to** N **do**

   repeat

   Generate $\boldsymbol{\theta}^{\star}$ from the prior distribution $\pi(\cdot)$

   Generate $\boldsymbol{z}$ from the likelihood $f(\cdot|\boldsymbol{\theta}^{\star})$

   until $\rho\{\nu(\boldsymbol{t}), \nu(\boldsymbol{z})\} < \epsilon$

   set $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} = \boldsymbol{\theta}^{\star}$

   end for

points $\boldsymbol{\theta}^{\star}$ is sampled from the prior distribution $\pi(\cdot)$. Given $\boldsymbol{\theta}^{\star}$, a data set $\boldsymbol{z}$ is then simulated under the statistical model in study. If the generated $\boldsymbol{z}$ is *too different* from the observed data $\boldsymbol{t}$, the sampled parameter value $\boldsymbol{\theta}^{\star}$ is discarded. In precise terms, the sampled parameter value is accepted with tolerance $\epsilon \geq 0$ if $\rho\{\nu(\boldsymbol{y}), \nu(\boldsymbol{z})\} < \epsilon$, where ν represents a set of summary statistics and $\rho\{\cdot,\cdot\}$ is a distance measure. A good choices of these quantities is essential to obtain the desired result, and therefore this step is the most complicated one when implementing the ABC. If the procedure is repeated a number of N times, the methods results in a sequence of *good* posterior estimates $\boldsymbol{\theta}^{\star}_{\boldsymbol{1}}, \boldsymbol{\theta}^{\star}_{\boldsymbol{2}}, \cdots$. from which quantities of interest can be easily obtained. According to Biau et al. (2015) the justification of the ABC algorithm lies in the fact that when ν is a sufficient statistic of $\boldsymbol{\theta}$, the distribution of $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ converges to the genuine posterior distribution when ϵ goes to zero. However, when the likelihood is unknown, then a sufficient statistics can hardly be identified. In that case, the recommendation of Fearnhead and Prangle (2012) is to find a low dimensional summary of the data that is sufficiently informative about $\boldsymbol{\theta}$. An alternative to select a value of ϵ is to keep the *best* estimates, that is to accept just a certain proportion of the sampled values (e.g. 5%, 1%, 0.5%), those closest to the real data according to the defined distance measure. Wilkinson (2013) describes ϵ as controlling a trade-off between computability and accuracy, because while smaller values of ϵ lead to better approximate of the true posterior, they also lead to lower acceptance rates and consequently, more computation must be done to get a given sample size.

For more details about the implementation of this method and possible extensions to it, we refer the reader to Csilléry et al. (2010) or Mengersen et al. (2013). Although to our knowledge it has not been used for *BMAP* estimation before, it has been recently suggested by Dean et al. (2014) for parametric estimation of hidden Markov models. In Chapter 4 we provide an illustration of this method in the context of a bivariate *MMPP* process.

1.4 Structure of this dissertation

This dissertation is structured in five chapters. In Chapter 1 the most important concepts and properties needed to develop our analyses was presented. Firstly, a brief review of the point processes and renewal processes was introduced. The Poisson process, phase type distribution and Markov processes were also described. Afterwards, the *BMAP* and *BMMPP* and their main properties were detailed, and finally a review of the main estimation procedures proposed in the literature for *BMMPP* and *BMAP* have been presented.

The results in Chapter 2 concern the identifiability of the stationary $BMMPP_m(K)$, where is proved that these processes are identifiable. This result extends the finding made by Ryden (1996), who proved the identifiability of the $MMPP_m$. In addition, some results related to the correlation and autocorrelation structure of the $BMMPP_m(K)$ are provided. The results of this chapter are published in Yera et al. (2019a).

Chapter 3, motivated by the identifiability of the $BMMPP_m(K)$ proved in the Chapter 2, focuses on exploring the capabilities of $BMMPP_m(K)$ in modeling real phenomena that occur in the form of point processes with group arrivals. First, it is proven that there is a set of moments that characterize the $BMMPP_2(K)$. This characterization leads to a sequential fitting approach via a moments matching method. We also provide a simulated and a real teletraffic data set to illustrate the performance of the novel fitting approach, which is compared to that of the EM algorithm. On the other hand, the queue length distributions at departures in the queueing system $BMMPP/M/1$ is also estimated. The results of this chapter are published in Yera et al. (2019b).

With the aim to extend the properties of $MMPP_2$, Chapter 4 presents the two-dimensional version of this process. The motivation for this extension arises from a real failures of trains data set, in which it could be appreciated that neither the times between failures, nor the distances travelled by the trains between failures were independent. In addition, these two variables were correlated. Therefore, we propose a novel two-dimensional *MMPP* that allows for dependence among the two sequences of inter-event magnitudes, while at the same time preserves the *MMPP* properties, marginally. Such generalization is based on the Marshall-Olkin exponential distribution. The identifiability of this novel process is proved and the inference is undertaken through a method combining a matching moments approach and an ABC algorithm. The performance of the method is shown on simulated and real dataset representing failures of a public transport company.

Finally, in Chapter 5 we summarize the most significant contributions of this dissertation, and also provide a short description of possible research lines.

References

Ali, S. and Pievatolo, A. (2016). High quality process monitoring using a class of inter-arrival time distributions of the renewal process. *Computers & Industrial Engineering*, 94:45 – 62.

Asmussen, S. (2008). *Applied probability and queues*, volume 51. Springer

Science & Business Media.

Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian journal of statistics*, 23:419–441.

Bean, N. and Green, D. (1999). When is a MAP poisson? *Mathematical and Computer Modelling*, 82:127–142.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beirlant, J., Teugels, J. L., and Vynckier, P. (1996). *Practical analysis of extreme values*, volume 50. Leuven University Press Leuven.

Biau, G., Cérou, F., and Guyader, A. (2015). New insights into approximate bayesian computation. In *Annales de l'IHP Probabilités et statistiques*, volume 51, pages 376–403.

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.

Breuer, L. (2002). An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, 112:123–138.

Buchholz, P. (2003). An EM-algorithm for MAP fitting from real traffic data. In *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 218–236. Springer.

Carrizosa, E. and Ramírez-Cobo, P. (2014). Maximum likelihood estimation in the two-state Markovian arrival process. *arXiv preprint arXiv:1401.3105*, Working paper.

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

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

Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution*, 25(7):410–418.

Dean, T. A., Singh, S. S., Jasra, A., and Peters, G. W. (2014). Parameter estimation for hidden markov models with intractable likelihoods. *Scandinavian Journal of Statistics*, 41(4):970–987.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Ephraim, Y. and Merhav, N. (2002). Hidden markov processes. *IEEE Transactions on information theory*, 48(6):1518–1569.

Ephraim, Y. and Roberts, W. J. (2008). An em algorithm for markov modulated markov processes. *IEEE Transactions on Signal Processing*, 57(2):463–470.

Eum, S., Harris, R., and Atov, I. (2007). A matching model for MAP-2 using moments of the counting process. In *Proceedings of the International Network Optimization Conference, INOC 2007*, Spa, Belgium.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: Semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Fearnhead, P. and Sherlock, C. (2006). An exact Gibbs sampler for the Markov modulated poisson process. *Journal of the Royal Statististical Society: Series B*, 65(5):767–784.

Green, D. (1998). *$MAP/PH/1$ departure processes*. PhD thesis, School of Applied Mathematics, University of Adelaide, South Australia.

He, Q.-M. (2014). *Fundamentals of matrix-analytic methods*, volume 365. Springer.

He, Q.-M. and Zhang, H. (2006). PH-invariant polytopes and coxian representations of phase type distributions. *Stochastic Models*, 22(3):383–409.

He, Q.-M. and Zhang, H. (2008). An algorithm for computing minimal coxian representations. *INFORMS Journal on Computing*, 20:179–190.

He, Q.-M. and Zhang, H. (2009). Coxian representations of generalized Erlang distributions. *Acta Mathematicae Applicatae Sinica. English Series*, 25:489–502.

Horváth, G. and Okamura, H. (2013). A fast EM algorithm for fitting marked Markovian arrival processes with a new special structure. In *European Workshop on Performance Engineering*, pages 119–133. Springer.

Horváth, M. and Telek, M. (2002). Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In *Performance evaluation of complex systems: Techniques and Tools, IFIP Performance 2002, in: LNCS Tutorial Series, vol. 2459*, pages 405–434.

Ito, H., Amari, S.-I., and Kobayashi, K. (1992). Identifiability of hidden markov information sources and their minimum degrees of freedom. *IEEE transactions on information theory*, 38(2):324–333.

Jatiningsih, L., Respatiwulan, Susanti, Y., Handayani, S. S., and Hartatik (2019). The parameter estimation of conditional intensity function temporal point process as renewal process using bayesian method and its application on the data of earthquake in east nusa tenggara. *Journal of Physics: Conference Series*, 1217:12–63.

Kang, S. H., Kim, Y. H., Sung, D. K., and Choi, B. D. (2002). An application of markovian arrival process (MAP) to modeling superposed atm cell streams. *IEEE Transactions on Communications*, 50(4):633–642.

Kim, C. S., Klimenok, V., Mushko, V., and Dudin, A. (2010). The bmap/ph/n retrial queueing system operating in markovian random environment. *Computers & Operations Research*, 37(7):1228–1237.

Klemm, A., Lindemann, C., and Lohmann, M. (2003). Modeling IP traffic using Batch Markovian Arrival Process. *Performance Evaluation*, 54(2):149–173.

Kriege, J. and Buchholz, P. (2011). Correlated phase-type distributed random numbers as input models for simulations. *Performance Evaluation*, 68(11):1247–1260.

Landriault, D. and Shi, T. (2015). Occupation times in the map risk model. *Insurance: Mathematics and Economics*, 60:75–82.

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM.

Li, S. and Ren, J. (2013). The maximum severity of ruin in a perturbed risk process with markovian arrivals. *Statistics & Probability Letters*, 83(4):993–998.

Lucantoni, D. (1991). New results for the single server queue with a Batch Markovian Arrival Process. *Stochastic Models*, 7:1–46.

Lveill, G. and Hamel, E. (2019). Compound trend renewal process with discounted claims: a unified approach. *Scandinavian Actuarial Journal*, 2019(3):228–246.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian Computational methods. *Statistics and Computing*, 22(6):1167–1180.

Marshall, A. H., Mitchell, H., and Zenga, M. (2015). *Modelling the Length of Stay of Geriatric Patients in Emilia Romagna Hospitals Using Coxian Phase-Type Distributions with Covariates*, pages 127–139. Springer International Publishing, Cham.

Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, 110(4):1321–1326.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2006). Reliability of a system under two types of failures using a Markovian arrival process. *Operations Research Letters*, 34:525–5530.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2014a). Matrix stochastic analysis of the maintainability of a machine under shocks. *Reliability Engineering & System Safety*, 121:11–17.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2014b). A redundant n-system under shocks and repairs following markovian arrival processes. *Reliability Engineering & System Safety*, 130:69–75.

Nakagawa, T. (2011). *Stochastic processes: With applications to reliability theory*. Springer Science & Business Media.

Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779.

Okamura, H. and Dohi, T. (2009). Faster maximum likelihood estimation algorithms for markovian arrival processes. In *2009 Sixth international conference on the quantitative evaluation of systems*, pages 73–82. IEEE.

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

Okamura, H., Kishikawa, H., and Dohi, T. (2013). Application of deterministic annealing em algorithm to map/ph parameter estimation. *Telecommunication Systems*, 54(1):79–90.

Pievatolo, A. and Ruggeri, F. (2010). Bayesian modelling of train door reliability. *The Oxford Handbook of Applied Bayesian Analysis. Oxford University Press, Oxford*, pages 271–294.

Pievatolo, A., Ruggeri, F., and Argiento, R. (2003). Bayesian analysis and prediction of failures in underground trains. *Quality and Reliability Engineering International*, 19(4):327–336.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798.

Ramírez-Cobo, P. and Lillo, R. (2012). New results about weakly equivalent MAP_2 and MAP_3 processes. *Methodology and Computing in Applied Probability*, 14(3):421–444.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010). Nonidentifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47(3):630–649.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2014a). Identifiability of the $MAP_2/G/1$ queueing system. *Top*, 22(1):274–289.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2017). Bayesian analysis of the stationary MAP_2. *Bayesian Analysis*, 12(4):1163–1194.

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

Riska, A., Diev, V., and Smirni, E. (2004). Efficient fitting of long-tailed data sets into phase-type distributions. *Performance Evaluation Journal*, 55:147–164.

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 and System Safety*, 134:126–133.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016a). Analytical issues regarding the lack of identifiability of the non-stationary MAP_2. *Performance Evaluation*, 102:1–20.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016b). Dependence patterns for modeling simultaneous events. *Reliability Engineering and System Safety*, 154:19–30.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016c). Nonidentifiability of the two-state $BMAP$. *Methodology and Computing in Applied Probability*, 18(1):81–106.

Rydén, T. (1996a). An EM algorithm for estimation in Markov modulated Poisson processes. *Computational Statistics & Data Analysis*, 21(4):431–447.

Rydén, T. (1996b). 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.

Scott, S. (1999). Bayesian analysis of the two state markov modulated poisson process. *Journal of Computational and Graphical Statistics*, 8(3):662–670.

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.

Telek, M. and Horváth, G. (2007). A minimal representation of markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153–1168.

Wang, K., Tao, M., Chen, W., and Guan, Q. (2015). Delay-aware energy-efficient communications over nakagami-m fading channel with $MMPP$ traffic. *IEEE Transactions on Communications*, 63(8):3008–3020.

Wilkinson, R. D. (2013). Approximate bayesian computation (abc) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, 12(2):129–141.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019a). Findings about the two-state BMMPP for modeling point processes in reliability and queueing systems. *Applied Stochastic Models in Business and Industry*, 35(2):177–190.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019b). Fitting procedure for the two-state Batch Markov modulated Poisson process. *European Journal of Operational Research*, 279(1):79–92.

Part II

Batch Markov modulated Poisson processes

CHAPTER 2

Findings about the BMMPP for modeling dependent and simultaneous data in reliability and queueing systems

Yoel G. Yera[1], Rosa E. Lillo[1,2] and Pepa Ramírez-Cobo[3,4]

[1] Department of Statistics, Universidad Carlos III de Madrid, Spain
[2] UC3M-BS Institute of Financial Big Data Getafe, Madrid, Spain
[3] Department of Statistics and Operations Research, Universidad de Cádiz, Spain
[4] IMUS, Institute of Mathematics, Univ. Sevilla, Spain

Abstract

The Batch Markov Modulated Poisson Process (*BMMPP*) is a subclass of the versatile Batch Markovian Arrival process (*BMAP*) which has been widely used for the modeling of dependent and correlated simultaneous events (as arrivals, failures or risk events). Both theoretical and applied aspects are examined in this paper. On one hand, the identifiability of the stationary $BMMPP_m(K)$ is proven, where K is the maximum batch size and m is the number of states of the underlying Markov chain. This is a powerful result for inferential issues. On the other hand, some novelties related to the correlation and autocorrelation structures are provided.

Keywords: Markov-modulated Poisson process (*MMPP*), Batch Markovian arrival process (*BMAP*), Correlation structure, Identifiability

1 Introduction

The Batch Markovian arrival process (noted *BMAP*, see Neuts Neuts (1979)) has been suggested in the literature for modeling dependent data representing the occurrence of an arrival, failure or risk event. The *BMAP*s constitute a large class of point processes that allows for non-exponential and dependent times between (possibly correlated) consecutive batch events. It is known that stationary $(B)MAP$s are capable of approximating any stationary (batch) point process (Asmussen and Koole, 1993), which suggests the versatility and range of applications of such processes. Therefore, different classes of *BMAP*s have been considered in a number of real life contexts where batch dependent occurrence times are commonly observed, as in queueing, teletraffic, reliability or insurance. See for example, Banerjee et al. (2015), Sikdar and Samanta (2016), Banik and Chaudhry (2016), Ghosh and Banik (2017), Montoro-Cazorla and Pérez-Ocón (2006), Montoro-Cazorla et al. (2009), Okamura et al. (2009), Kang et al. (2002), Casale et al. (2010), Wu et al. (2011), (Ramírez-Cobo et al. (2014a), Ramírez-Cobo et al. (2014b)), Montoro-Cazorla and Pérez-Ocón (2015) and Liu et al. (2015). The single arrival *BMAP*, the *MAP*, and the general *BMAP* are highly parametrized models where, in practice, only inter-event times and batch sizes are usually observed. Therefore, such processes commonly suffer from identifiability problems, which occur when different representations lead to the same likelihood function for the observed data. The study of the identifiability is crucial when estimation of the model parameters is to be considered. In particular, the non-identifiability of a process has serious negative consequences: the likelihood function may be highly multimodal, implying that standard methods (as the EM algorithm) will be strongly dependent on the starting values, running the risk of getting stuck at a poor local maximum.

Different works have dealt with the problem of identifiability in *BMAP*-related processes, especially for the *MAP* and some of its subclasses as the well-known Markov-modulated Poisson process (*MMPP*). See for example, Rydén (1996), where it is proven the identifiability of the *MMPP*. On the other hand, Bodrog et al. (2008) provided a canonical and unique representation for the MAP_2. Another example is Ramírez-Cobo et al. (2010), where it is shown that the MAP_2 is not identifiable. Furthermore, the conditions under which two different sets of parameters induce identical stationary laws for the observable process are given. Also Ramírez-Cobo and Lillo (2012) partially solved the identifiability problem for the stationary MAP_3. For the non-stationary MAP_2 the lack of identifiability is studied in Rodríguez et al. (2016a). On the other hand, Rodríguez et al. (2016c) proves the non-identifiability of the stationary $BMAP_2$ noted $BMAP_2(K)$, where K is the maximum batch size. For the case where events occur simultaneously, Rodríguez et al. (2016c) seems to be the unique paper devoted to study the identifiability issue, up to the authors knowledge.

As commented previously, the *MMPP* is an identifiable class of *MAP*, a fact that has eased its statistical estimation, see Landon et al. (2013), Özekici and Soyer (2006), Özekici and Soyer (2003), Fearnhead and Sherlock (2006), Heyman and Lucantoni (2003), Scott and Smyth (2003) and Scott (1999). In this paper we consider the batch counterpart of the *MMPP*, the so-called Batch Markov-modulated Poisson process, noted *BMMPP*. This process has been already considered in the literature (Dudin, 1998; Akar and Sohraby, 2009; Revzina, 2010; Takine, 2016) for modeling real-time multimedia communication systems and computer networks systems. However, in most of such papers, a reduced version of the *BMMPP* with batch probabilities independent from the states of the underlying Markov chain, is used. The *BMAP* version of this kind of process has been mentioned as a *MAP* with i.i.d. batch arrivals by (Lucantoni, 1991, 1993). Here, we consider the general *BMMPP*, for which two major contributions are provided. We study the versatility of the process for modeling correlated batch events through the autocorrelation function of the batch sizes. Our findings show the suitability of the $BMMPP_2$ for fitting positively correlated batch sizes. Second, with the future aim of carrying out statistical inference of the process, we prove the identifiability of the *BMMPP*.

This paper is structured as follows. The *BMMPP* is introduced in Section 2. Section 3 provides new results concerning the autocorrelation function of the batch sizes for the $BMMPP_2(K)$. In section 4 the identifiability of the general $BMMPP_m(K)$, where m denotes the number of states of the underlying Markov chain, is proven. Finally, Section 5 is devoted to summarize the conclusions and some extensions of this work.

2 Description of the stationary $\boldsymbol{BMMPP}_m(K)$

In this section, the Batch Markov-Modulated Poisson Process, noted $BMMPP_m(K)$, where K is the maximum batch size and m is the number of states of the underlying Markov chain, is formally defined. Also, some properties that will be used throughout this paper are reviewed.

The $BMMPP_m(K)$ is a Poisson process whose rate is modulated by an exogenous, irreducible Markov process, $\{J(t) : t \geq 0\}$, with state space $S := \{s_1, s_2, ..., s_m\}$, a generator matrix Q and an initial distribution $\boldsymbol{\alpha}$. Whenever $J(t) = s_i$, an event occur according to a Poisson process with rate λ_i ($\lambda_i > 0$), and this status remains unchanged while the process remains in this state. As soon as J enters another state, $s_j \in S$, the arrival Poisson process alters accordingly. The process behaves as follows: at the end of an exponentially distributed sojourn time in state i, with mean $1/\lambda_i$, two possible state transitions can occur. First, with probability p_{ij0}, no event occurs and the system enters into a different state $j \neq i$. Second, with probability p_{iik}, a batch event of size k is produced, if the state of the process is s_i, and the system continues in the same state i.

A $BMMPP_m(K)$ can be represented by the set of rate matrices $\mathcal{B} = (\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K})$ such that

$$\begin{aligned}
(\boldsymbol{D_0})_{ii} = &\quad -\lambda_i && 1 \leq i \leq m \\
(\boldsymbol{D_0})_{ij} = &\quad \lambda_i p_{ij0} && 1 \leq i, j \leq m \quad i \neq j \\
(\boldsymbol{D_k})_{iik} = &\quad \lambda_i p_{iik} && 1 \leq i \leq m \quad 1 \leq k \leq K, \\
(\boldsymbol{D_k})_{ijk} = &\quad 0 && 1 \leq i, j \leq m \quad i \neq j,
\end{aligned} \tag{2.1}$$

where

$$\sum_{j=1, j\neq i}^{m} p_{ij0} + \sum_{k=1}^{K} p_{iik} = 1 \quad \text{for all } i = \{1, 2\}.$$

The definition of the rate matrices implies that

$$\boldsymbol{Q} = \sum_{k=0}^{K} \boldsymbol{D_k}$$

is the infinitesimal generator of the underlying Markov process $J(t)$, with stationary probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_m)$, satisfying $\boldsymbol{\pi} Q = 0$ and $\boldsymbol{\pi e} = 1$, where $\boldsymbol{e}$ is a vector of ones.

Figure 2.1 illustrates a realization of the $BMMPP_2(K)$, where the dashed line corresponds to transitions without events and the solid lines correspond to transitions where an event of size $b_i \in \{1, \ldots, K\}$ occurs.

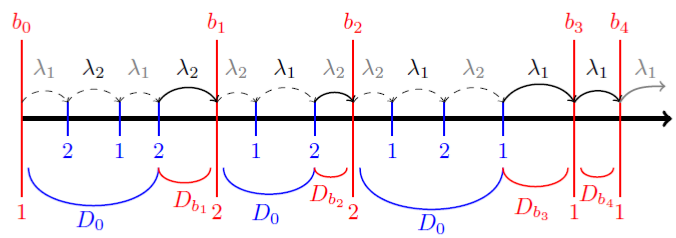


Figure 2.1: Transition diagram for the $BMMPP_2(K)$. The dashed line corresponds to transitions without events, governed by $\boldsymbol{D_0}$, and the solid lines correspond to transitions of size b_k, governed by $\boldsymbol{D_{b_k}}$.

2.1 Performance measures

A review of the performance measures regarding the $BMMPP_m(K)$ is given next. If S_n denotes the state of the $BMMP_m(K)$ at the time of the n-th event, B_n the batch size of that event and T_n the time between the $(n-1)$-th and n-th events, then the process $\{S_{n-1}, \sum_{i=1}^{n} Ti, B_n\}_{n=1}^{\infty}$, is a Markov renewal process (see for example, Chakravarthy (2010)). Furthermore, if

$$\boldsymbol{D} = \sum_{k=1}^{K} \boldsymbol{D_k},$$

then $\{S_n\}_{n=0}^{\infty}$ is a Markov chain with transition probability matrix

$$\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D}.$$

On the other hand, the variables T_ns are phase-type distributed with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}\}$, such that $\boldsymbol{\phi}$ is the stationary probability vector associated to $\boldsymbol{P^*}$, computed as $\boldsymbol{\phi} = (\boldsymbol{\pi D e})^{-1}\boldsymbol{\pi D}$ (see Chakravarthy (2010) and Latouche and Ramaswami (1999)). In consequence, the moments of T_n in the stationary case are given by

$$\mu_r = E(T^r) = r!\boldsymbol{\phi}(-\boldsymbol{D_0})^{-r}\boldsymbol{e}, \quad \text{for } r \geq 1, \tag{2.2}$$

and the auto-correlation function is

$$\rho_T(l) = \rho(T_1, T_{l+1}) = \frac{\mu\boldsymbol{\pi}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D}]^l(-\boldsymbol{D_0})^{-1}\boldsymbol{e} - \mu_1^2}{\sigma_T^2},$$

where $\sigma^2 = \mu_2 - \mu_1^2$.

According to Rodríguez et al. (2016c), the mass probability function of the stationary batch size, B, is

$$P(B = k) = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D^k e}, \quad \text{for } k = 1, ..., K,$$

from which the moments of B are obtained as

$$\beta_r = E[B^r] = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_r^* e}, \quad \text{for } r \geq 1, \tag{2.3}$$

where $\boldsymbol{D_r^*} = \sum_{k=1}^{K} k^r \boldsymbol{D_k}$. Also, the autocorrelation function in the stationary version of the process $\rho_B(l)$ is given by

$$\rho_B(l) = \rho(B_1, B_{l+1}) = \frac{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D}]^{l-1}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e} - \beta_1^2}{\sigma_B^2}, \tag{2.4}$$

where β_1 and $\sigma_B^2 = \beta_2 - \beta_1^2$ are computed from (2.3).

Finally, in Rodríguez et al. (2016c) it is proven that the Laplace-Stieltjes transform (LST) of the n first inter-event times and batch sizes of a stationary $BMAP_m(K)$ is given by

$$f^*_{T,B}(\boldsymbol{s},\boldsymbol{z}) = \boldsymbol{\phi}(s_1\boldsymbol{I} - \boldsymbol{D_0})^{-1}\boldsymbol{\xi}(z_1)...(s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1}\boldsymbol{\xi}(z_n)\boldsymbol{e}, \tag{2.5}$$

where $\boldsymbol{s} = (s_1, ..., s_n)$, $\boldsymbol{z} = (z_1, ..., z_n)$ and $\boldsymbol{\xi}(z_i) = \sum_{k=1}^{K} \boldsymbol{D_k} z_i^k$, for $i = 1, ..., n$.

3 The autocorrelation function of the batch size for the two-state *BMMPP*

It usually has been assumed in applications of the *BMMPP* that $p_{iik} = p_{jjk}$ for all i, j; or, in other words, the state of the underlying Markov process and the size events are independent, see Cordeiro and Kharoufeh (2011), Takine (2016), Revzina (2010) and Dudin (1998). This kind of process is a particular case of the *MAP* with i.i.d. batch arrivals introduced by (Lucantoni, 1991, 1993). This definition can be restrictive in practice since both the autocorrelation function of the batch size as well as the correlation coefficient between the batch size and the times between the occurrence of events are zero. This results is an immediate consequence of the definition of the process. That implies independence between the arrival process and the batch process.

In order to avoid the limited behavior of this simplified version of the *BMMPP* we consider from now the general $BMMPP_2(K)$ with probabilities p_{iik} dependent on state $i \in \{1, 2\}$. Figures 2.2 and 2.3 show the first-lag autocorrelation coefficient of the batch size and the correlation coefficient between the inter-event times and the batch sizes, respectively for a sequence of 10000 randomly simulated $BMMPP_2(2)$s. Since the autocorrelation function of the batch size decreases with the lag value (see, Rodríguez et al. (2016a)), then, it can be deduced from the figures that the autocorrelation function of the batch size and the correlation coefficient between the batch size and the inter-events arrival times are not zero in the general case.

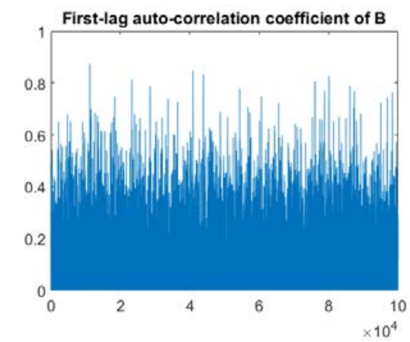


Figure 2.2: Values of the first-lag autocorrelation coefficient of the batch size for a total 100000 simulated $BMMPP_2(2)$ with the general formulation.

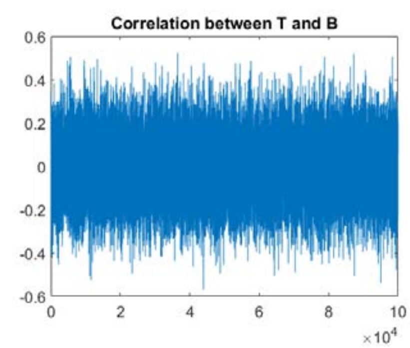


Figure 2.3: Values of the correlation between the batch size and the inter-event time for a total 100000 simulated $BMMPP_2(2)$ with the general formulation.

The auto-correlation function of the batch size is crucial when the modeling capability of the process is of interest. The auto-correlation function for the inter-event times, ρ_T, is the same for the $BMAP_2(K)$ and MAP_2. Then, the results obtained for the MAP_2 by Heindl et al. (2006) and Ramírez-Cobo and Carrizosa (2012) also apply for the $BMMPP_2(K)$. In particular, it is known that the lag-one auto-correlation coefficient, $\rho_T(1)$, is upper-bounded by 0.5 and the auto correlation function is exponentially decreasing in absolute value. On the other hand, Kang and Sung (1995) prove that for any $MMPP_2$, $\rho_T(l) \geq 0$ for all l.

In the case of the event sizes, Rodríguez et al. (2016b) give a characterization of the auto-correlation functions in terms of the eigenvalues of the stochastic matrix $\boldsymbol{P^*}$. For the two-states process, this representation allows to prove that the autocorrelation function decreases geometrically to zero through four different patterns. Moreover, through simulation, it is shown that the function values lie in [-1,1]. In Figure 2.2 it can be seen how the first-lag auto-correlation coefficient of the batch sizes for the $BMMPP_2(2)$ may also take values very close to 1, but negative values are not obtained.

In this section it is proven that the autocorrelation function of the batch sizes of the $BMMPP_2(K)$, $\rho_B(l)$ as in (2.4), is non-negative.

Lemma 1. *Consider a* $\mathrm{BMMPP}_2(2)$ *and let* $\rho_B(1)$ *denote the first-lag autocorrelation coefficient of the batch sizes. Then,* $\rho_B(1) \geq 0$.

Proof. A stationary $BMMPP_2(2)$ is represented by $\mathcal{B} = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ where

$$\boldsymbol{D_0} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{120} \\ \lambda_2 p_{210} & -\lambda_2 \end{pmatrix} = \begin{pmatrix} x & y \\ r & u \end{pmatrix},$$

$$\boldsymbol{D_1} = \begin{pmatrix} p_{111}\lambda_1 & 0 \\ 0 & p_{221}\lambda_2 \end{pmatrix} = \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix},$$

$$\boldsymbol{D_2} = \begin{pmatrix} p_{112}\lambda_1 & 0 \\ 0 & p_{222}\lambda_2 \end{pmatrix} = \begin{pmatrix} -x-y-w & 0 \\ 0 & -r-u-q \end{pmatrix}.$$

Consider the first-lag autocorrelation coefficient, $\rho_B(1)$. It is not difficult to see that after some computations, the numerator in (2.4), taking $l = 1$

and $K = 2$, becomes

$$\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}[\boldsymbol{I} - \boldsymbol{e}\boldsymbol{\phi}](-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e} = \frac{ry[w(u+r) - q(x+y)]^2}{(xu - ry)(rx + ry + ry + yu)^2}.$$

Similarly, the denominator in (2.4) is found as

$$\begin{aligned}\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_2^*}\boldsymbol{e} - (\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e})^2 =& \frac{(rw + yq)[r(-x - y - w)]}{(rx + ry + ry + yu)^2} \\ &+ \frac{(rw + yq)[y(-r - u - q)]}{(rx + ry + ry + yu)^2}.\end{aligned}$$

Therefore,

$$\rho_B(1) = \frac{ry[w(u+r) - q(x+y)]^2}{(xu - ry)(rw + yq)[r(-x - y - w) + y(-r - u - q)]}. \tag{2.6}$$

Since,

$$\begin{aligned} ry &= \lambda_2 p_{210} \lambda_1 p_{120} \geq 0, \\ rw + yq &= \lambda_2 p_{210} \lambda_1 p_{111} + \lambda_1 p_{120} \lambda_2 p_{221} \geq 0, \\ r(-x - y - w) + y(-r - u - q) &= \lambda_2 p_{210} \lambda_1 p_{112} + \lambda_1 p_{120} \lambda_2 p_{222} \geq 0, \\ xu - ry &= Det(D_0) = \lambda_1 \lambda_2 (1 - p_{120} p_{210}) \geq 0, \end{aligned}$$

then, it can be concluded that $\rho_B(1)$ as in (2.6) satisfies $\rho_B(1) \geq 0$. □

Lemma 1 is extended to Lemma 2 for the case $K \geq 3$.

Lemma 2. *Consider a* BMMPP$_2(K)$ *with* $K \geq 3$ *and let* $\rho_B(1)$ *denote the first-lag autocorrelation coefficient of the batch sizes. Then,* $\rho_B(1) \geq 0$.

Proof. In this case, the process will be represented by $\mathcal{B} = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \ldots, \boldsymbol{D_K}\}$ where

$$\boldsymbol{D_0} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{120} \\ \lambda_2 p_{210} & -\lambda_2 \end{pmatrix} = \begin{pmatrix} x & y \\ r & u \end{pmatrix},$$

$$\boldsymbol{D_k} = \begin{pmatrix} p_{11k}\lambda_1 & 0 \\ 0 & p_{22k}\lambda_2 \end{pmatrix} = \begin{pmatrix} w_k & 0 \\ 0 & q_k \end{pmatrix} \quad \text{for } 1 \le k \le K-1,$$

$$\begin{aligned} \boldsymbol{D_K} &= \begin{pmatrix} p_{11K}\lambda_1 & 0 \\ 0 & p_{22K}\lambda_2 \end{pmatrix} \\ &= \begin{pmatrix} -x-y-\sum_{k=1}^{K-1} w_k & 0 \\ 0 & -r-u-\sum_{k=1}^{K-1} q_k \end{pmatrix}. \end{aligned}$$

Here, the numerator in (2.4) with $l = 1$ and general K turns out

$$\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}[\boldsymbol{I}-\boldsymbol{e}\boldsymbol{\phi}](-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e} = \frac{ry[W_1(u+r)-Q_1(x+y)]^2}{(xu-ry)(rx+ry+ry+yu)^2}, \tag{2.7}$$

where $W_1 = \sum_{k=1}^{K-1}(K-k)w_k$ and $Q_1 = \sum_{k=1}^{K-1}(K-k)q_k$. Since

$$\begin{aligned} ry &= \lambda_2 p_{210}\lambda_1 p_{120} \ge 0, \\ xu - ry &= Det(\boldsymbol{D_0}) = \lambda_1\lambda_2(1-p_{120}p_{210}) \ge 0, \end{aligned}$$

then Eq. (2.7) is positive or equal to zero. Similarly, the denominator is given by

$$\begin{aligned} \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}[\boldsymbol{D_2^*}-\boldsymbol{D_1^*}\boldsymbol{e}\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}]\boldsymbol{e} &= \frac{[r(W_2-KW_1)](rx+2ry+yu)}{(rx+ry+ry+yu)^2} \\ &\quad + \frac{[y(Q_2-KQ_1)](rx+2ry+yu)}{(rx+ry+ry+yu)^2} \\ &\quad - \frac{(rW_1+yQ_1)^2}{(rx+ry+ry+yu)^2} \end{aligned} \tag{2.8}$$

where $W_2 = \sum_{k=1}^{K-1} k(K-k)w_k$ and $Q_2 = \sum_{k=1}^{K-1} k(K-k)q_k$. We prove next the non-negativity of (2.8). First, define:

$$H_K = [r(W_2-KW_1)+y(Q_2-KQ_1)]\,(rx+2ry+yu)-(rW_1+yQ_1)^2$$

$$
\begin{aligned}
= & -\left(\sum_{k=1}^{K-1}(K-k)^2[rw_k+yq_k]\right)(rx+2ry+yu) \\
& -\left(\sum_{k=1}^{K-1}(K-k)[rw_k+yq_k]\right)^2 .
\end{aligned}
\tag{2.9}
$$

It can be seen that when $K=3$, expression (2.9) reduces to:

$$
\begin{aligned}
H_3 = \; & [2(rw_1+yq_1)+(rw_2+yq_2)][-r(x+y+w_1+w_2)] \\
& +[2(rw_1+yq_1)+(rw_2+yq_2)][-y(r+u+q_1+q_2)] \\
& +[rw_1+yq_1][-r(x+y+w_1+w_2)-r(x+y+w_1)] \\
& +[rw_1+yq_1][-y(r+u+q_1+q_2)-y(r+u+q_1)] \\
\geq \; & 0,
\end{aligned}
$$

We now proceed by induction. Assume that $H_{K-1}\geq 0$, then after some calculations H_K can be rewritten as

$$
\begin{aligned}
H_K = \; & \left(\sum_{k=1}^{K-1}(K-k)[rw_k+yq_k]\right)\left[-r(x+y+\sum_{k=1}^{K-1}w_k)\right] \\
& \left(\sum_{k=1}^{K-1}(K-k)[rw_k+yq_k]\right)\left[-y(r+u+\sum_{k=1}^{K-1}q_k)\right] \\
& +\left(\sum_{k=1}^{K-2}[(K-k-1)][rw_k+yq_k]\right)\left[-r(x+y+\sum_{k=1}^{K-1}w_k)\right] \\
& +\left(\sum_{k=1}^{K-2}[(K-k-1)][rw_k+yq_k]\right)\left[-y(u+r+\sum_{k=1}^{K-1}q_k)\right] \\
& -\left(\sum_{k=1}^{K-2}(K-k-1)^2[rw_k+yq_k]\right)(rx+2ry+yu) \\
& -\left(\sum_{k=1}^{K-2}(K-k-1)[rw_k+yq_k]\right)^2 \\
= \; & \left(\sum_{k=1}^{K-1}(K-k)[rw_k+yq_k]\right)[rw_K+yq_K]+H_{K-1}
\end{aligned}
$$

$$
\begin{aligned}
& + \left(\sum_{k=1}^{K-2} [(K-k-1)][rw_k + yq_k] \right) [rw_K + yq_K] \\
\geq \ & 0,
\end{aligned}
$$

since r, y, w_k, q_k are non-negative for all k and $H_{K-1} \geq 0$ by the induction hypothesis. Therefore, Eq. (2.8) is is positive or equal to zero and consequently $\rho_B(1) \geq 0$ is proven. □

Proposition 3. *Consider a* BMMPP$_2(K)$*, with autocorrelation function of the batch sizes given by* $\rho_B(l)$*, as in (2.4). Then,* $\rho_B(l) \geq 0$ *for all* $l \geq 1$*.*

Proof. In Rodríguez et al. (2016b) it is proven that the autocorrelation function of the batch sizes in a $BMAP_2(K)$ is given by

$$\rho_B(l) = \rho_B(1) q_B^{l-1}, \tag{2.10}$$

where q_B is the only eigenvalue of $\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D}$ less than 1 in absolute value. Note that, in the $BMMPP_2(K)$, $\boldsymbol{D}$ can be computed as

$$\boldsymbol{D} = \sum_{k=1}^{K} \boldsymbol{D_k} = \begin{pmatrix} -x-y & 0 \\ 0 & -r-u \end{pmatrix}.$$

Therefore, in this specific case, q_B is given by

$$q_B = \frac{(-x-y)(r+u)}{ry - xu}, \tag{2.11}$$

for all K. Since

$$
\begin{aligned}
r + u &= -\lambda_2(1 - p_{210}) \leq 0 \\
-x - y &= \lambda_1(1 - p_{120}) \geq 0 \\
xu - ry &= Det(\boldsymbol{D_0}) = \lambda_1\lambda_2(1 - p_{120}p_{210}) \geq 0,
\end{aligned}
$$

then, it can be concluded that q_B as in (2.11) satisfies $q_B \geq 0$ and consequently, from (2.10) and Lemmas 1-2, $\rho_B(l) \geq 0$ for all $l \geq 1$.

□

4 Identifiability of the $\boldsymbol{BMMPP}_m(K)$

Identifiability problems occur when different representations of the process lead to the same likelihood functions for the observable data. In order to develop an estimation method to fit real datasets to the model, a detailed examination of the identifiability of the process is critical. It is well known that the *MAP* and *BMAP* processes suffer from identifiability problems, but, on the other hand, in Rydén (1996), the identifiability of the *MMPP* was proven. Here, we extend such result to the $BMMPP_m(K)$ case. First, consider the definition of identifiability that we will use in the paper. It is valid to be adapted to real data sets.

Definition 8. *Let $\mathcal{B}$ be a representation of a* $\mathrm{BMMPP}_2(K)$ *and let T_n and B_n denote the time between the $(n-1)$-th and n-th event occurrences, and the batch size of the n-th event occurrence, respectively. Then $\mathcal{B}$ is said to be identifiable if there does not exist a different parametrization $\widetilde{\mathcal{B}}$, such that*

$$(T_1, ..., T_n, B_1, ..., B_n) \stackrel{d}{=} (\widetilde{T}_1, ..., \widetilde{T}_n, \widetilde{B}_1, ..., \widetilde{B}_n), \quad \textit{for all } n \geq 0,$$

where $\tilde{T}_i$ and $\tilde{B}_i$ are defined in analogous way as T_i and B_i, and where $\stackrel{d}{=}$ denotes equality in distribution.

In what follows we will concentrate on the LST, given in (2.5), in order to prove the identifiability of $BMMPP_2(K)$. Note that the equality in distribution is equivalent to the equality of the LSTs, $f^*_{T,B}(\boldsymbol{s}, \boldsymbol{z}) = f^*_{\tilde{T},\tilde{B}}(\boldsymbol{s}, \boldsymbol{z})$, for all $\boldsymbol{s}, \boldsymbol{z}$. First, we review the concept of permutation matrix and some of its properties that are useful to obtain the main result (for more details, see for example Horn and Johnson (1990)).

Definition 9. *A square matrix $\boldsymbol{P}$ is a permutation matrix if exactly one entry in each row and column is equal to 1 and all other entries are 0.*

Some properties concerning permutations matrices are:

P1. $\boldsymbol{PA}$ implies a permutations of the rows of $\boldsymbol{A}$, where $\boldsymbol{A}$ is an $m \times n$ matrix.

P2. $\boldsymbol{AP}$ permutes the columns of $\boldsymbol{A}$.

P3. A permutation matrix is orthogonal ($\boldsymbol{P}^{-1} = \boldsymbol{P}^T$)

P4. The permutation matrices are closed under product.

Next result establishes how to obtain equivalent representations to a given $BMMPP_m(K)$, using permutation matrices. Note that in this context of $BMMPP_m(K)$s, the multiplication by a permutation implies a change of the states' labels, but the process is the same. The only practical implication of this in modeling is that some order among the parameters have to be established to avoid switching from one representation to the equivalent one.

Lemma 3. *Let* $\mathcal{B} = \{\boldsymbol{D_0}, ..., \boldsymbol{D_K}\}$ *be a representation of a* $\mathrm{BMMPP}_m(K)$ *and let* $\mathcal{B}_P = \{\boldsymbol{PD_0P},\ \boldsymbol{PD_1P}, ...,\ \boldsymbol{PD_KP}\}$ *be a different representation where* P *is a permutation matrix. Then,* $\mathcal{B}$ *and* $\mathcal{B}_P$ *are equivalent representations of the same process.*

Proof. The equivalence between $\mathcal{B}$ and $\mathcal{B}_P$ is given by proving the equality of their respective LTs as in (2.5). Consider first $\boldsymbol{\pi_P}$, the stationary probability vector related to representation $\mathcal{B}_P$, which satisfies $\boldsymbol{\pi_P Q_P} = \boldsymbol{0}$. It is not difficult to see that

$$\begin{aligned} \boldsymbol{\pi_P Q_P} &= \boldsymbol{\pi_P} \left(\sum_{k=0}^{K} \boldsymbol{PD_kP} \right) \\ &= \boldsymbol{\pi_P P} \left(\sum_{k=0}^{K} \boldsymbol{D_k} \right) \boldsymbol{P} \\ &= \boldsymbol{\pi_P PQP} \end{aligned}$$

and

$$\boldsymbol{\pi_P PQP} = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{\pi_P PQ} = \boldsymbol{0}.$$

But $\boldsymbol{\pi Q} = \mathbf{0}$ and $\boldsymbol{\pi PPQ} = \boldsymbol{\pi Q}$, since by applying the same permutation in the columns of vector $\boldsymbol{\pi}$ ($\boldsymbol{\pi P}$) that in the in the rows of the matrix $\boldsymbol{Q}$ ($\boldsymbol{PQ}$) the product is not affected, therefore $\boldsymbol{\pi_P} = \boldsymbol{\pi P}$. On the other hand, let $\boldsymbol{\phi_P}$ denote the stationary probability vector with transitions events related to representation $\mathcal{B}_P$. Then

$$\begin{aligned}
\boldsymbol{\phi_P} &= \left[\boldsymbol{\pi_P}\left(\sum_{k=1}^{K}\boldsymbol{PD_kP}\right)\boldsymbol{e}\right]^{-1}\boldsymbol{\pi_P}\left(\sum_{k=1}^{K}\boldsymbol{PD_kP}\right)\boldsymbol{e} \\
&= \left[\boldsymbol{\pi PP}\left(\sum_{k=1}^{K}\boldsymbol{D_k}\right)\boldsymbol{Pe}\right]^{-1}\boldsymbol{\pi PP}\left(\sum_{k=1}^{K}\boldsymbol{D_k}\right)\boldsymbol{P} \\
&= [\boldsymbol{\pi PPDPe}]^{-1}\boldsymbol{\pi PPDP} \\
&= [\boldsymbol{\pi De}]^{-1}\boldsymbol{\pi DP} \\
&= \boldsymbol{\phi P}. \qquad (2.12)
\end{aligned}$$

From, the fact that $\boldsymbol{Pe} = \boldsymbol{e}$, the property P3 of permutation matrices and (2.12), we have that

$$\begin{aligned}
f^*_{T_P,B_P}(\boldsymbol{s},\boldsymbol{z}) &= \boldsymbol{\phi_P}\left[\prod_{i=1}^{n}(s_i\boldsymbol{I}-\boldsymbol{PD_0P})^{-1}\left(\sum_{l=1}^{k}\boldsymbol{PD_lP}z_i^l\right)\right]\boldsymbol{e} \\
&= \boldsymbol{\phi P}\left[\prod_{i=1}^{n}(\boldsymbol{P}s_i\boldsymbol{IP}-\boldsymbol{PD_0P})^{-1}\left(\sum_{l=1}^{k}\boldsymbol{PD_lP}z_i^l\right)\right]\boldsymbol{e} \\
&= \boldsymbol{\phi P}\left[\prod_{i=1}^{n}[\boldsymbol{P}(s_i\boldsymbol{I}-\boldsymbol{D_0})\boldsymbol{P}]^{-1}\boldsymbol{P}\left(\sum_{l=1}^{k}\boldsymbol{D_l}z_i^l\right)\boldsymbol{P}\right]\boldsymbol{e} \\
&= \boldsymbol{\phi P}\left[\prod_{i=1}^{n}\boldsymbol{P}^T(s_i\boldsymbol{I}-\boldsymbol{D_0})^{-1}\boldsymbol{P}^T\boldsymbol{P}\left(\sum_{l=1}^{k}\boldsymbol{D_l}z_i^l\right)\boldsymbol{P}\right]\boldsymbol{e} \\
&= \boldsymbol{\phi PP}^T\left[\prod_{i=1}^{n}(s_i\boldsymbol{I}-\boldsymbol{D_0})^{-1}\left(\sum_{l=1}^{k}\boldsymbol{D_l}z_i^l\right)\right]\boldsymbol{Pe} \\
&= \boldsymbol{\phi}\left[\prod_{i=1}^{n}(s_i\boldsymbol{I}-\boldsymbol{D_0})^{-1}\left(\sum_{l=1}^{k}\boldsymbol{D_l}z_i^l\right)\right]\boldsymbol{e} = f^*_{T,B}(\boldsymbol{s},\boldsymbol{z}),
\end{aligned}$$

and the lemma is proven. □

The next result is a direct consequence of the identifiability of the *MMPP*s.

Lemma 4. *Let* $\mathcal{B} = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_k}\}$ *and* $\tilde{\mathcal{B}} = \left\{\tilde{D}_0, \tilde{D}_1, ..., \tilde{D}_k\right\}$ *be two different but equivalent representations of a* $\mathrm{BMMPP}_m(K)$*. Then,* $\boldsymbol{D_0} = \boldsymbol{\tilde{D}_0}$ *and* $\boldsymbol{\Lambda} = \boldsymbol{\tilde{\Lambda}}$*, except by permutation, where* $\boldsymbol{\Lambda}$ *(*$\boldsymbol{\tilde{\Lambda}}$*) is the vector of exponential rates of* $\mathcal{B}$ *(*$\tilde{\mathcal{B}}$*).*

Proof. It is clear that if representations $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are equivalent, then, the *MMPP*s representations $\mathcal{B}' = (\boldsymbol{D_0}, \boldsymbol{D})$ and $\tilde{\mathcal{B}}' = \{\boldsymbol{\tilde{D}_0}, \boldsymbol{\tilde{D}}\}$ will be also equivalent, where $\boldsymbol{D} = \boldsymbol{D_1} + \ldots + \boldsymbol{D_K}$ and $\boldsymbol{\tilde{D}} = \boldsymbol{\tilde{D}_1} + ... + \boldsymbol{\tilde{D}_K}$. From the identifiability of the *MMPP*, $\boldsymbol{D_0} = \boldsymbol{\tilde{D}_0}$ and $\boldsymbol{D} = \boldsymbol{\tilde{D}}$ except by permutation. Hence, $\boldsymbol{\Lambda} = \boldsymbol{\tilde{\Lambda}}$, except by permutation. □

The main contribution of this section is the Theorem 1 that proves the identifiability of the general $BMMPP_m(K)$. However, some auxiliary results are previously necessary.

Proposition 4. *The* $\mathrm{BMMPP}_2(2)$ *is identifiable except by permutation.*

Proof. Consider two stationary $BMMPP_2(2)$s represented by $\mathcal{B} = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ and $\boldsymbol{\tilde{B}} = \left\{\boldsymbol{\tilde{D}_0}, \boldsymbol{\tilde{D}_1}, \boldsymbol{\tilde{D}_2}\right\}$. For $n = 1$, the LST of the inter-event times and batch sizes corresponding to B is given by

$$\begin{aligned}
f^*_{T,B}(s,z) &= \boldsymbol{\phi}(s\boldsymbol{I} - \boldsymbol{D_0})^{-1}(z\boldsymbol{D_1} + z^2\boldsymbol{D_2})\boldsymbol{e} \\
&= \boldsymbol{\phi}\frac{z}{(s-x)(s-u)-ry}\begin{pmatrix} s-u & y \\ r & s-x \end{pmatrix}\begin{pmatrix} w \\ q \end{pmatrix} \\
&\quad +\boldsymbol{\phi}\frac{z^2}{(s-x)(s-u)-ry}\begin{pmatrix} s-u & y \\ r & s-x \end{pmatrix}\begin{pmatrix} -x-y-w \\ -r-u-q \end{pmatrix} \\
&= \frac{z\left[\phi(sw-uw)+(1-\phi)rw+\phi yq+(1-\phi)(sq-xq)\right]}{(s-x)(s-u)-ry}
\end{aligned}$$

$$
\begin{aligned}
& +\frac{z^2\left[\phi(u-s)(x+y+w)+(1-\phi)r(-x-y-w)\right]}{(s-x)(s-u)-ry} \\
& +\frac{z^2\left[\phi y(-r-u-q)+(1-\phi)(x-s)(r+u+q)\right]}{(s-x)(s-u)-ry} \\
= \quad & \frac{z(s\alpha+\beta)+z^2(s\gamma-\beta+\eta)}{s^2+s\nu+\eta}
\end{aligned} \tag{2.13}
$$

where

$$
\begin{aligned}
\alpha &= \phi(w-q)+q \\
\beta &= \phi(-uw-rw+yq+xq)+(rw-xq) \\
\gamma &= \phi(r+u+q-x-y-w)-(r+u+q) \\
\eta &= xu-ry \\
\nu &= -x-u,
\end{aligned}
$$

and similarly for $\tilde{\mathcal{B}}$. If $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are equivalent, then from Lemma 4, $\boldsymbol{D_0} = \tilde{\boldsymbol{D}}_{\boldsymbol{0}}$, and therefore $x = \tilde{x}$, $y = \tilde{y}$, $u = \tilde{u}$, $r = \tilde{r}$, and consequently, $\nu = \tilde{\nu}$ and $\eta = \tilde{\eta}$. This implies that the equality of LSTs,

$$f^*_{T,B}(s,z) = f^*_{\tilde{T},\tilde{B}}(s,z), \qquad \text{for all } s, z.$$

becomes

$$z(s\alpha+\beta)+z^2(s\gamma-\beta) = z(s\tilde{\alpha}+\tilde{\beta})+z^2(s\tilde{\gamma}-\tilde{\beta}), \qquad \text{for all } s, z. \tag{2.14}$$

Substituting first $s = 0$ and $z = 2$ in (2.14), and later from $s = 1$ and $z = -1$, leads to $\beta = \tilde{\beta}$ and $\alpha = \tilde{\alpha}$:

$$
\begin{aligned}
\beta &= \phi(-uw-rw+yq+xq)+(rw-xq) \\
&= \phi(-u\tilde{w}-r\tilde{w}+y\tilde{q}+x\tilde{q})+(r\tilde{w}-x\tilde{q}) \\
&= \tilde{\beta}
\end{aligned} \tag{2.15}
$$

Similarly,

$$\alpha = \phi(w-q)+q = \phi(\tilde{w}-\tilde{q})+\tilde{q} = \tilde{\alpha}$$

from which

$$\tilde{w} = \frac{\phi(w - q + \tilde{q}) + q - \tilde{q}}{\phi}. \tag{2.16}$$

If (2.16) is substituted in (2.15), then

$$\begin{aligned}
\phi[-w(u+r) + q(y+x)] + (rw - xq) = &- \phi \left[\frac{\phi(w - q + \tilde{q}) + q - \tilde{q}}{\phi}\right](u+r) \\
&+ \phi(y\tilde{q} + x\tilde{q}) + r\frac{\phi(w - q + \tilde{q}) + q - \tilde{q}}{\phi} \\
&- x\tilde{q} \\
= &- [\phi(-q + \tilde{q}) + q - \tilde{q}](u+r) \\
&- \phi w(u+r) \\
&+ \phi(y\tilde{q} + x\tilde{q}) + r\frac{\phi(-q + \tilde{q}) + q - \tilde{q}}{\phi} \\
&- x\tilde{q} + rw,
\end{aligned}$$

hence

$$\begin{aligned}
\phi(yq + xq) - xq &= -[\phi(-q + \tilde{q}) + q - \tilde{q}](u+r) \\
&\quad +\phi(y\tilde{q} + x\tilde{q}) + \left(r\frac{\phi(-q + \tilde{q}) + q - \tilde{q}}{\phi} - x\tilde{q}\right), \\
&= (1-\phi)(\tilde{q} - q)(u+r) \\
&\quad +\phi\tilde{q}(y+x) + \left(\frac{r}{\phi}(1-\phi)(q - \tilde{q}) - x\tilde{q}\right), \\
&= \left(u + r - \frac{r}{\phi}\right)(1-\phi)(\tilde{q} - q) \\
&\quad +\tilde{q}(\phi(y+x) - x).
\end{aligned} \tag{2.17}$$

From (2.17) it can be concluded that $q = \tilde{q}$. Therefore, from (2.16) $\tilde{w} = w$ and consequently, $\boldsymbol{D_1} = \tilde{\boldsymbol{D}}_{\boldsymbol{1}}$ and $\boldsymbol{D_2} = \tilde{\boldsymbol{D}}_{\boldsymbol{2}}$. □

Proposition 4 proves that the stationary $BMMPP_2(2)$ is an identifiable process. The next Proposition goes further and ensures the identifiability of the stationary $BMMPP_m(2)$, for all $m \geq 3$.

Proposition 5. *The* BMMPP$_m$(2) *is identifiable except by permutation.*

Proof. We proceed by induction in m. The initial case ($m = 2$) was proven by Proposition 4. We assume as hypothesis of induction that if two $BMMPP_{m-1}(K)$ are equivalent then its transition matrices are respectively equal. Consider two equivalent $BMMPP_m(2)$s given by $\mathcal{B}_m(2) = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ and $\tilde{\mathcal{B}}_m(2) = \left\{\tilde{\boldsymbol{D}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}, \tilde{\boldsymbol{D}}_{\boldsymbol{2}}\right\}$ and obtain from them the $BMMPP_{m-1}(2)$ representations $\mathcal{B}^m_{m-1}(2) = \{\boldsymbol{D^m_0}, \boldsymbol{D^m_1}, \boldsymbol{D^m_2}\}$ and $\tilde{\mathcal{B}}^m_{m-1}(2) = \{\tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{1}}, \tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{2}}\}$, where $\boldsymbol{D^m_0}, \boldsymbol{D^m_1}, \boldsymbol{D^m_2}$ are $m-1 \times m-1$ matrices given by

$$\boldsymbol{D^m_i} = \begin{pmatrix} d_{11i} & \dots & d_{1m-2i} & d_{1m-1i} + d_{1mi} \\ \vdots & \ddots & \vdots & \vdots \\ d_{m-21i} & \dots & d_{m-2m-2i} & d_{m-2m-1i} + d_{m-2mi} \\ d_{m-11i} + d_{m1i} & \dots & d_{m-1m-2i} + d_{mm-2i} & d_{m-1m-1i} + d_{m-1mi} + d_{mm-1i} + d_{mmi} \end{pmatrix}$$

with $\boldsymbol{D_i} = (d_{lki})$. Similarly for $\tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{1}}, \tilde{\boldsymbol{D}}^{\boldsymbol{m}}_{\boldsymbol{2}}$.

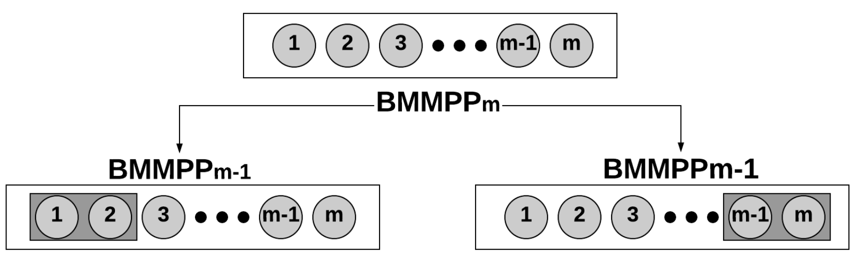


Figure 2.4: This Diagram represents the grouping of the states in the construction of $\mathcal{B}^1_{m-1}(2)$ (left) $\mathcal{B}^m_{m-1}(2)$ (right) from $\mathcal{B}_m(2)$ (top).

Note that $\mathcal{B}^m_{m-1}(2)$ and $\tilde{\mathcal{B}}^m_{m-1}(2)$ are derived from $\mathcal{B}_m(2)$ and $\tilde{\mathcal{B}}_m(2)$ respectively grouping the m-th and $m-1$-th states in a single state (see Figure 2.4 for clarification). With this construction the inter-events arrival times and the batch size in $\mathcal{B}^m_{m-1}(2)$ and $\mathcal{B}_m(2)$ are the same and similarly for $\tilde{\mathcal{B}}^m_{m-1}(2)$ and $\tilde{\mathcal{B}}_m(2)$. According to Definition 8 the identifiability of the process depends only of the inter-events arrival times and the batch sizes, therefore from the equivalence of $\mathcal{B}_m(2)$ and $\tilde{\mathcal{B}}_m(2)$, the equivalence of

$\mathcal{B}_{m-1}^{m}(2)$ and $\tilde{\mathcal{B}}_{m-1}^{m}(2)$ is obtained. Now, from the hypothesis induction for $m-1$

$$\lambda_i p_{iik} = \lambda_i \tilde{p}_{iik} \quad \text{for } 1 \leq i \leq m-2 \text{ and } k = 1, 2. \tag{2.18}$$

Similarly, consider $\mathcal{B}_{m-1}^{1}(2) = \{\boldsymbol{D_0^1}, \boldsymbol{D_1^1}, \boldsymbol{D_2^1}\}$ and $\tilde{\mathcal{B}}_{m-1}^{1}(2) = \{\tilde{\boldsymbol{D}}_{\boldsymbol{0}}^{\boldsymbol{1}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}^{\boldsymbol{1}}, \tilde{\boldsymbol{D}}_{\boldsymbol{2}}^{\boldsymbol{1}}\}$, where $\boldsymbol{D_0^1}, \boldsymbol{D_1^1}, \boldsymbol{D_2^1}$ are also $(m-1) \times (m-1)$ matrices given by

$$\boldsymbol{D_i^1} = \begin{pmatrix} d_{11i} + d_{12i} + d_{21i} + d_{22i} & d_{13i} + d_{23i} & \dots & d_{1mi} + d_{2mi} \\ d_{31i} + d_{32i} & d_{33i} & \dots & d_{3mi} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1i} + d_{m2i} & d_{3mi} & \dots & d_{mmi} \end{pmatrix},$$

with $\boldsymbol{D_i} = (d_{lki})$. Similarly for $\tilde{\boldsymbol{D}}_{\boldsymbol{0}}^{\boldsymbol{1}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}^{\boldsymbol{1}}, \tilde{\boldsymbol{D}}_{\boldsymbol{2}}^{\boldsymbol{1}}$. Again, $\mathcal{B}_{m-1}^{1}(2)$ and $\tilde{\mathcal{B}}_{m-1}^{1}(2)$ are equivalent, which can be proven using the same reasoning as with $\mathcal{B}_{m-1}^{m}(2)$ and $\tilde{\mathcal{B}}_{m-1}^{m}(2)$ and the hypothesis of induction. Therefore,

$$\lambda_i p_{iik} = \lambda_i \tilde{p}_{iik} \quad \text{for } 3 \leq i \leq m \text{ and } k = 1, 2. \tag{2.19}$$

From (2.18) and (2.19), $\boldsymbol{D_k} = \tilde{\boldsymbol{D}}_{\boldsymbol{k}}$, for all k, which completes the proof. □

Next Theorem extends the previous result, proving the identifiability of the stationary $BMMPP_m(K)$, for all $K \geq 3$.

Theorem 1. *The* BMMPP$_m(K)$ *is identifiable except by permutation.*

Proof. We proceed by induction in K. The initial case ($K = 2$) was proven by Proposition 5. Consider two equivalent $BMMPP_m(K)$s given by $\mathcal{B}_m(K) = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_k}\}$ and $\tilde{\mathcal{B}}_m(K) = \left\{\tilde{\boldsymbol{D}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}, ..., \tilde{\boldsymbol{D}}_{\boldsymbol{k}}\right\}$ and generate from them two $BMMPP_m(K-1)$ equivalent representations $\mathcal{B}_m^K(K-1) = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_{K-2}}, \boldsymbol{D_{K-1}} + \boldsymbol{D_K}\}$ and $\tilde{\mathcal{B}}_m^K(K-1) = \left\{\tilde{\boldsymbol{D}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}, ..., \tilde{\boldsymbol{D}}_{\boldsymbol{K-2}}, \tilde{\boldsymbol{D}}_{\boldsymbol{K-1}} + \tilde{\boldsymbol{D}}_{\boldsymbol{K}}\right\}$. For the proof of the equivalence between $\mathcal{B}_m^K(K-1)$ and $\tilde{\mathcal{B}}_m^K(K-1)$ see Appendix A. Then, from the induction hypothesis,

$$\boldsymbol{D_{K-1}} + \boldsymbol{D_K} = \boldsymbol{\tilde{D}_{K-1}} + \boldsymbol{\tilde{D}_K} \text{ and } \boldsymbol{D_k = \tilde{D}_k} \quad \text{for } 1 \leq k \leq K-2. \quad (2.20)$$

Similarly, consider $\mathcal{B}_m^1(K-1) = \{\boldsymbol{D_0}, \boldsymbol{D_1} + \boldsymbol{D_2}, \boldsymbol{D_3}, ..., \boldsymbol{D_K}\}$ and $\tilde{\mathcal{B}}_m^1(K-1)\left\{\boldsymbol{\tilde{D}_0}, \boldsymbol{\tilde{D}_1} + \boldsymbol{\tilde{D}_2}, \boldsymbol{\tilde{D}_3}, ..., \boldsymbol{\tilde{D}_K}\right\}$. Again, These two $BMMPP_m(K)$ are equivalent and from the hypothesis of induction

$$\boldsymbol{D_1} + \boldsymbol{D_2} = \boldsymbol{\tilde{D}_1} + \boldsymbol{\tilde{D}_2} \quad \text{and} \quad \boldsymbol{D_k} = \boldsymbol{\tilde{D}_k}, \quad \text{for } 3 \leq k \leq K. \qquad (2.21)$$

From (2.20) and (2.21), $\boldsymbol{D_k} = \boldsymbol{\tilde{D}_k}$, for all k, which completes the proof. □

5 Conclusions

This paper considers the batch counterpart of the well-known Markov-Modulated Poisson Process, the $BMMPP_m(K)$, a point process of interest in real-life contexts as reliability or queueing, since it allows for the modeling of dependent inter-event times and dependent batch sizes. Two main problems concerning the *BMMPP* have been addressed. On one hand, we prove the identifiability of the $BMMMPP_m(K)$, a property inherited from that of the *MMPP*. The identifiability of the process is of crucial importance when inference is to be considered, as we plan to do as future work. Either from a Bayesian viewpoint as in Scott (1999) or Ramírez-Cobo et al. (2017), or from a moments matching method as in Rodríguez et al. (2015), a statistical inference approach may be defined for fitting real datasets. On the other hand, the non-negativity of the autocorrelation function of the batch sizes of the $BMMPP_2(K)$ is proven. This property makes the process suitable when positively correlated batch sizes are observed. Finally, an important extension to this work would be to derive theoretical properties concerning correlation bounds for both the inter-event times and batch size autocorrelation functions when $m \geq 3$. Work on this issues is underway.

Acknowledgments

Research partially supported by research grants and projects MTM2015-65915-R and ECO2015-66593-P (Ministerio de Economía y Competitividad, Spain) and P11-FQM-7603, FQM-329 (Junta de Andalucía, Spain) and Fundación BBVA. The authors thank the anonymous referees whose comments and suggestions improved the understanding of this paper.

Appendix A: Proof of the equivalence between $\mathcal{B}_m^K(K-1)$ and $\tilde{\mathcal{B}}_m^K(K-1)$

Since $\mathcal{B}_K = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K}\}$ and $\tilde{\mathcal{B}}_K = \{\tilde{\boldsymbol{D}}_{\boldsymbol{0}}, \tilde{\boldsymbol{D}}_{\boldsymbol{1}}, ..., \tilde{\boldsymbol{D}}_{\boldsymbol{K}}\}$ are equivalent, then $\boldsymbol{D_0} = \tilde{\boldsymbol{D}}_{\boldsymbol{0}}$ and

$$f^*_{T,B}(s,z) = f^*_{\tilde{T},\tilde{B}}(s,z), \quad \text{for all } s, z,$$

or equivalently,

$$\boldsymbol{\phi} S(s_1, ..., s_n, z_1, ..., z_n)\boldsymbol{e} = \boldsymbol{\phi}\tilde{S}(s_1, ..., s_n, z_1, ..., z_n)\boldsymbol{e} \quad \text{for all } s, z, \tag{2.22}$$

where

$$S(s_1, ..., s_n, z_1, ..., z_n) = (s_1\boldsymbol{I} - \boldsymbol{D_0})^{-1} \sum_{k_1=1}^{K} \boldsymbol{D_{k_1}} z_1^{k_1} ... (s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1} \sum_{k_n=1}^{K} \boldsymbol{D_{k_n}} z_n^{k_n}$$

and similarly for $\tilde{S}$. Equality (2.22) can be rewritten as

$$\begin{aligned} &\boldsymbol{\phi} S(s_1, ..., s_{n-1}, z_1, ..., z_{n-1}) \times \\ &\times (s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1} \sum_{k_n=1}^{K} \boldsymbol{D_{k_n}} z_n^{k_n}\boldsymbol{e} \quad = \quad \boldsymbol{\phi}\tilde{S}(s_1, ..., s_{n-1}, z_1, ..., z_{n-1}) \times \\ &\qquad\qquad\qquad\qquad\qquad \times (s_n\boldsymbol{I} - \boldsymbol{D_0})^{-1} \sum_{k_n=1}^{K} \tilde{\boldsymbol{D}}_{\boldsymbol{k_n}} z_n^{k_n}\boldsymbol{e}. \end{aligned} \tag{2.23}$$

Next, consider the following three block of calculations related to value z_n:

Step1: Compute K times in both sides of (2.23) the partial derivative with respect to z_n:

$$\begin{aligned}\boldsymbol{\phi} S(s_1,...,s_{n-1},z_1,...,z_{n-1}) \times & \\ \times(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1}\boldsymbol{D_K}K!\boldsymbol{e} &= \boldsymbol{\phi}\tilde{S}(s_1,...,s_{n-1},z_1,...,z_{n-1}) \times \\ &\quad \times(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1}\tilde{\boldsymbol{D}}_{\boldsymbol{K}}K!\boldsymbol{e}. \qquad (2.24)\end{aligned}$$

Step 2: Multiply (2.24) by $(z_n^{K-1}-z_n^K)/K!$:

$$\begin{aligned}\boldsymbol{\phi} S(s_1,...,s_{n-1},z_1,...,z_{n-1}) \times & \qquad (2.25) \\ \times(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1}\boldsymbol{D_K}(z_n^{K-1}-z_n^K)!\boldsymbol{e} &= \boldsymbol{\phi}\tilde{S}(s_1,...,s_{n-1},z_1,...,z_{n-1}) \times \\ &\quad \times(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1}\tilde{\boldsymbol{D}}_{\boldsymbol{K}}(z_n^{K-1}-z_n^K)\boldsymbol{e}.\end{aligned}$$

Step 3: Summing (2.23) + (2.25), we arrive to

$$\boldsymbol{\phi} S^*(s_1,...,s_n,z_1,...,z_n)\boldsymbol{e} = \boldsymbol{\phi}\tilde{S}^*(s_1,...,s_n,z_1,...,z_n)\boldsymbol{e}, \qquad (2.26)$$

where

$$\begin{aligned}S^*(s_1,...,s_n,z_1,...,z_n) =& \boldsymbol{\phi} S(s_1,...,s_{n-1},z_1,...,z_{n-1})(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1} \\ & \times\left[\sum_{k_n=1}^{K-1}\boldsymbol{D}_{\boldsymbol{k_n}}z_n^{k_n}+\boldsymbol{D_K}z_n^{K-1}\right]\boldsymbol{e}\end{aligned}$$

and equivalently

$$\begin{aligned}\tilde{S}^*(s_1,...,s_n,z_1,...,z_n) =& \boldsymbol{\phi}\tilde{S}(s_1,...,s_{n-1},z_1,...,z_{n-1})(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1} \\ & \times\left[\sum_{k_n=1}^{K-1}\tilde{\boldsymbol{D}}_{\boldsymbol{k_n}}z_n^{k_n}+\tilde{\boldsymbol{D}}_{\boldsymbol{K}}z_n^{K-1}\right]\boldsymbol{e}.\end{aligned}$$

If the previous Steps 1-3 are reproduced for $z_{n-1}, z_{n-2}, \ldots, z_1$, the equivalency between $\mathcal{B}^2_{K-1}$ and $\tilde{\mathcal{B}}^2_{K-1}$ is obtained. Since the LTS for $\mathcal{B}^2_{K-1}$ is

$$\begin{aligned} f^*_{T,B}(s,z) &= \boldsymbol{\phi}(s_1\boldsymbol{I}-\boldsymbol{D_0})^{-1}\left[\sum_{k_1=1}^{K-1}\boldsymbol{D_{k_1}}z_1^{k_1}+\boldsymbol{D_K}z_1^{K-1}\right]\times \\ &\quad ...\times(s_n\boldsymbol{I}-\boldsymbol{D_0})^{-1}\left[\sum_{k_n=1}^{K-1}\boldsymbol{D_{k_n}}z_1^{k_n}+\boldsymbol{D_K}z_n^{K-1}\right]\boldsymbol{e} \end{aligned}$$

and similarly for $\tilde{\mathcal{B}}^2_{K-1}$.

Finally, by a parallel procedure, the equivalence between $\mathcal{B}^1_{K-1}$ and $\tilde{\mathcal{B}}^1_{K-1}$ is also derived.

References

Akar, N. and Sohraby, K. (2009). System-theoretical algorithmic solution to waiting times in semi-markov queues. *Performance Evaluation*, 66(11):587–606.

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

Banerjee, A., Gupta, U., 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.

Banik, A. and Chaudhry, M. (2016). Efficient computational analysis of stationary probabilities for the queueing system $BMAP/G/1/N$ with or without vacation (s). *INFORMS Journal on Computing*, 29(1):140–151.

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.

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

Chakravarthy, S. R. (2010). Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science*.

Cordeiro, J. D. and Kharoufeh, J. P. (2011). Batch markovian arrival processes (bmap). *Wiley Encyclopedia of Operations Research and Management Science*.

Dudin, A. (1998). Optimal multithreshold control for a $BMAP/G/1$ queue with n service modes. *Queueing Systems*, 30(3-4):273–287.

Fearnhead, P. and Sherlock, C. (2006). An exact Gibbs sampler for the Markov modulated poisson process. *Journal of the Royal Statististical Society: Series B*, 65(5):767–784.

Ghosh, S. and Banik, A. (2017). An algorithmic analysis of the $BMAP/MSP/1$ generalized processor-sharing queue. *Computers & Operations Research*, 79:1–11.

Heindl, A., Mitchell, K., and van de Liefvoort, A. (2006). Correlation bounds for second-order MAPs with application to queueing network decomposition. *Performance Evaluation*, 63(6):553–577.

Heyman, D. and Lucantoni, D. (2003). Modeling multiple IP traffic streams with rate limits. *IEEE/ACM Transactions on Networking*, 11(6):948–958.

Horn, R. A. and Johnson, C. R. (1990). *Matrix analysis*. Cambridge university press.

Kang, S. H., Kim, Y. H., Sung, D. K., and Choi, B. D. (2002). An application of markovian arrival process (MAP) to modeling superposed atm cell streams. *IEEE Transactions on Communications*, 50(4):633–642.

Kang, S. H. and Sung, D. K. (1995). Two-state $MMPP$ modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times. In *Proceedings of GLOBECOM'95*, volume 2, pages 1422–1426. IEEE.

Landon, J., Özekici, S., and Soyer, R. (2013). A markov modulated poisson model for software reliability. *European Journal of Operational Research*, 229(2):404–410.

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM.

Liu, B., Cui, L., Wen, Y., and Shen, J. (2015). A cold standby repairable system with working vacations and vacation interruption following markovian arrival process. *Reliability Engineering & System Safety*, 142:1–8.

Lucantoni, D. (1991). New results for the single server queue with a Batch Markovian Arrival Process. *Stochastic Models*, 7:1–46.

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

Montoro-Cazorla, D. and Pérez-Ocón, R. (2006). Reliability of a system under two types of failures using a Markovian arrival process. *Operations Research Letters*, 34:525–5530.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2015). A reliability system under cumulative shocks governed by a bmap. *Applied Mathematical Modelling*, 39(23-24):7620–7629.

Montoro-Cazorla, D., Pérez-Ocón, R., and Segovia, M. (2009). Replacement policy in a system under shocks following a Markovian arrival process. *Reliability Engineering and System Safety*, 94:497–502.

Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779.

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

Özekici, S. and Soyer, R. (2003). Reliability of software with an operational profile. *European Journal of Operational Research*, 149(2):459–474.

Özekici, S. and Soyer, R. (2006). Semi-markov modulated poisson process: probabilistic and statistical analysis. *Mathematical Methods of Operations Research*, 64(1):125–144.

Ramírez-Cobo, P. and Carrizosa, E. (2012). A note on the dependence structure of the two-state Markovian arrival process. *Journal of Applied Probability*, 49:295–302.

Ramírez-Cobo, P. and Lillo, R. (2012). New results about weakly equivalent MAP_2 and MAP_3 processes. *Methodology and Computing in Applied Probability*, 14(3):421–444.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010). Nonidentifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47(3):630–649.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2014a). Identifiability of the $MAP_2/G/1$ queueing system. *Top*, 22(1):274–289.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2017). Bayesian analysis of the stationary MAP_2. *Bayesian Analysis*, 12(4):1163–1194.

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

Revzina, E. (2010). Stochastic models of data flows in the telecommunication networks. *Computer Modelling and New Technologies*, 14(2):29–34.

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 and System Safety*, 134:126–133.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016a). Analytical issues regarding the lack of identifiability of the non-stationary MAP_2. *Performance Evaluation*, 102:1–20.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016b). Dependence patterns for modeling simultaneous events. *Reliability Engineering and System Safety*, 154:19–30.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016c). Nonidentifiability of the two-state $BMAP$. *Methodology and Computing in Applied Probability*, 18(1):81–106.

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.

Scott, S. (1999). Bayesian analysis of the two state markov modulated poisson process. *Journal of Computational and Graphical Statistics*, 8(3):662–670.

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.

Sikdar, K. and Samanta, S. (2016). Analysis of a finite buffer variable batch service queue with batch markovian arrival process and servers vacation. *Opsearch*, 53(3):553–583.

Takine, T. (2016). Analysis and computation of the stationary distribution in a special class of markov chains of level-dependent $M/G/1$-type and its application to $BMAP/M/\infty$ and $BMAP/M/c+M$ queues. *Queueing Systems*, 84(1-2):49–77.

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*, 35:1184–1193.

CHAPTER 3

Fitting procedure for the two-state Batch Markov modulated Poisson process

Yoel G. Yera[1], Rosa E. Lillo[1,2] and Pepa Ramírez-Cobo[3,4]

[1] Department of Statistics, Universidad Carlos III de Madrid, Spain
[2] UC3M-BS Institute of Financial Big Data Getafe, Madrid, Spain
[3] Department of Statistics and Operations Research, Universidad de Cádiz, Spain
[4] IMUS, Institute of Mathematics, Univ. Sevilla, Spain

Abstract

The Batch Markov Modulated Poisson Process (*BMMPP*) is a subclass of the versatile Batch Markovian Arrival Process (*BMAP*) which has been proposed for the modeling of dependent events occurring in batches (such as group arrivals, failures or risk events). This paper focuses on exploring the possibilities of the *BMMPP* for the modeling of real phenomena involving point processes with group arrivals. The first result in this sense is the characterization of the $BMMPP_2(K)$ by a set of moments related to the inter-event time and batch size distributions. This characterization leads to a sequential fitting approach via a moments matching method. The performance of the novel fitting approach is illustrated on both simulated and a real teletraffic data set, and compared to that of the EM algorithm. In addition, as an extension of the inference approach, the queue length distributions at departures in the queueing system $BMMPP/M/1$ is also estimated.

Keywords: Stochastic processes. Markov modulated Poisson process (*MMPP*). Moments matching method. Teletraffic data.

1 Introduction

In this work, we propose a fitting approach for correlated times between the occurrence of events (that may occur in batches) via a general subclass of the Batch Markovian Arrival Process (*BMAP*), the Batch Markov Modulated Poisson Process (*BMMPP*). Events can be understood from multiple contexts: failures in an electronic system, arrivals of packets of bytes in a teletraffic setting, arrivals of customers in a queue or risk events, among others. The *BMAP* constitutes a large class of point processes that allows for non-exponential and dependent times between the occurrence of events, which may occur in batches (that is, more than one event at a time). *BMAP*s were first introduced by Neuts (1979), although the current and more tractable description is due to Lucantoni (1991). It is known that stationary *BMAP*s are capable of approximating any stationary batch point process (Asmussen and Koole, 1993) which points to the versatility of the process. In addition, the *BMAP* is a tractable process from an analytical viewpoint, since most of the associated descriptors and probabilities of interest can be computed in a straightforward way. For these reasons, *BMAP*s have been widely considered in a number of real-life contexts, such as queueing, teletraffic, reliability, hydrology or insurance, where dependent events (possibly occurring in batches) are commonly observed. For a recent account of the literature on *BMAP*s applications, we refer the reader to Ramírez-Cobo et al. (2014), Banerjee et al. (2015), Liu et al. (2015), Montoro-Cazorla and Pérez-Ocón (2015), Singh et al. (2016), Sikdar and Samanta (2016), Banik and Chaudhry (2016), Ghosh and Banik (2017), and Buchholz and Kriege (2017).

The complexity and versatility of *BMAP*s increase with the number of parameters defining the process, which is related to the identifiability issue. In the context of *BMAP*s, the lack of identifiability may be formulated along the lines of Rydén (1996b) or Ramírez-Cobo et al. (2010). Specifically, if T_n

and B_n represent the time between the $(n-1)$-th and n-th events ocurrences, and the batch size of the n-th event in a *BMAP* noted by $\mathcal{B}$, then $\mathcal{B}$ is said to be non-identifiable if there exists a differently parametrized *BMAP*, noted as $\widetilde{\mathcal{B}}$, such that

$$(T_1, \ldots, T_n, B_1, \ldots, B_n) \stackrel{d}{=} \left(\widetilde{T}_1, \ldots, \widetilde{T}_n, \widetilde{B}_1, \ldots, \widetilde{B}_n\right) \quad \text{for all } n \geq 1,$$

where $\stackrel{d}{=}$ denotes equality of joint distributions, and $\widetilde{T}_n$ and $\widetilde{B}_n$ represent the inter-event times and batch sizes of the *BMAP* noted as $\widetilde{\mathcal{B}}$. The lack of a unique representation affects the statistical inference of the process: if the process is non-identifiable, this means that the likelihood function of the inter-event times will be multimodal, and therefore any likelihood-based fitting algorithm will turn out to be strongly dependent on the starting point. Because of this, the issue of identifiability has been broadly studied in the literature for certain classes of *BMAP*s, see for instance Green (1998); Bean and Green (1999); He and Zhang (2006, 2008, 2009); Rydén (1996b); Ramírez-Cobo et al. (2010); Ramírez-Cobo and Lillo (2012); Rodríguez et al. (2016a,c); Yera et al. (2019). As a result, it is known that both the Markov modulated Poisson process (*MMPP*) (Heffes and Lucantoni, 1986; Scott, 1999; Scott and Smyth, 2003; Fearnhead and Sherlock, 2006; Landon et al., 2013) and its batch counterpart, the *BMMPP* considered in this paper, are identifiable.

Taking advantage of the identifiability of the *BMMPP*, this paper addresses the problem of statistical inference for the $BMMPP_2(K)$ where K represents the maximum batch size. The choice of the $BMMPP_2(K)$ over higher order $BMMPP_m(K)$s (that is, processes with $m \geq 3$) is motivated by several reasons. First, the model considered is characterized by a smaller number of parameters, a fact that facilitates the estimation process. Second, it is expected that higher order $BMMPP_m(K)$s present more versatility and are able to model more complex patterns (Rodríguez et al. (2016b) give some empirical results in this line); however, to our knowledge there are no studies exploring in depth such degrees of versatility and in consequence, it is impossible to know *a priori* which is the smallest order m needed for fitting a given data set. Finally, as will be shown in Section 3, the $BMMPP_2(K)$ can be completely characterized in terms of a set of $2(K+1)$ moments related to the inter-event times and batch size distribution, which naturally leads to a moments-matching fitting approach. However, this characterization in terms of moments remains an open question for the case $m \geq 3$, and will be the subject of future work as indicated in the conclusions section.

The contribution of this paper is two-fold. On one hand and as commented before, it is proven that the $BMMPP_2(K)$, which is represented by $2(K+1)$ parameters, is characterized by a set of $2(K+1)$ moments concerning the distributions of both the inter-event times and batch sizes. On the other hand, a sequential estimation approach for fitting real data sets is derived and illustrated for simulated and real data sets. At this point, some remarks concerning statistical inference for the $BMAP$s need to be made. First, concerning the observed information, in most papers it is assumed that the sequence of inter-event times, $\boldsymbol{t} = (t_1, t_2, ..., t_n)$ (and if it is the case, of batch sizes $\boldsymbol{b} = (b_1, b_2, ..., b_n)$) constitute the available observed samples. This implies that many components of the process (such as the transition times or sequence of visited states) remain unobserved. Other authors instead consider that the observed information is related to the counting process (number of accumulated events at some time instants, for example), see Andersen and Nielsen (2002); Arts (2017); Nasr et al. (2018). In this paper, the approach considered will be the first one. Second, there are a number of papers in the literature dealing with strategies (either Bayesian, frequentist or moments matching based) for estimation of some types of MAPs (characterized by single events at a time). In these works, either the MAPs considered are identifiable (as the $MMPP$, see Rydén (1994); Rydén (1996a); Scott (1999)) or non-identifiable, but with a known canonical form (such as the MAP_2 see Eum et al. (2007); Bodrog et al. (2008); Carrizosa and Ramírez-Cobo (2014); Ramírez-Cobo et al. (2017)). However, if events occurring in batches are observed, fewer studies dealing with inference for the $BMAP$ can be found and to our knowledge they all are based on the EM algorithm, see for example Breuer (2002); Klemm et al. (2003). In this paper the performance of the proposed sequential fitting algorithm shall be compared to that of the EM, as designed in such papers.

The paper is structured as follows. After a brief review of the $BMMPP_2(K)$ in Section 2, the moments characterization for the $BMMPP_2(K)$ is proven in Section 3. Section 3.2 analyzes in depth the case $K = 2$ and Section 3.3 extends the findings for the case $K \geq 3$. The characterization in terms of moments leads to the sequential fitting algorithm presented and illustrated in Section 4. After the detailed description of the method in Section 4.1, its performance on simulated traces is illustrated in Section 4.2 and a comparison with the EM algorithm is provided in Section 4.3. Finally, Section 4.4 considers a real application of the novel approach: the modeling of a well-referenced data set from the teletraffic context. In the numerical analyses, the estimation of the queue length distribution at departures in a $BMMPP/M/1$ queueing system is also considered. Finally, Section 5 presents conclusions and delineates possible directions for future research.

2 Description of the stationary $BMMPP_2(K)$

In this section, the $BMMPP_2(K)$, where K is the maximum batch size, is formally defined. Also, some properties that will be used throughout this paper are reviewed. Consider a two-state Markov process $J(t)$ with generator Q on $\{1, 2\}$. For each state $i \in \{1, 2\}$, events occur according to a Poisson process with rate λ_i and each event has a batch distribution on $\{1, \ldots, K\}$ that also depends on $J(t)$. In other words, whenever $J(t) = i$, it is said that the process is in state i at time t and this status remains unchanged while the process remains in this state. As soon as the Markov process enters another state j ($j \in \{1, 2\}$), then the Poisson process alters accordingly. Specifically, the $BMMPP_2(K)$ behaves as follows: at the end of an exponentially distributed sojourn time in state i, with mean $1/\lambda_i$, two possible state transitions can occur. First, with probability p_{ij0}, no event occurs and the system enters into a different state $j \neq i$. Second, with probability p_{iik}, an event of batch size k is produced if the state of the process is i, and the system continues in the same state. It is clear that

$$p_{ij0} + \sum_{k=1}^{K} p_{iik} = 1 \qquad i, j = 1, 2, \quad i \neq j.$$

A $BMMPP_2(K)$ can be thus expressed in terms of the initial probability vector and the parameters $\{\boldsymbol{\lambda}, \boldsymbol{P_0}, .., \boldsymbol{P_K}\}$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, and $\boldsymbol{P_0}$, $\boldsymbol{P_1}$,...,$\boldsymbol{P_K}$ are 2×2 transition probability matrices with (i, j)-th elements p_{ijk}, for $k = 1, ..., K$. On the other hand, instead of transition probability matrices, any $BMMPP_2(K)$ can also be characterized in terms of rate (or intensity) matrices. In the case of the $BMMPP_2(K)$, these rate matrices are $\{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K}\}$ where

$$\begin{aligned}
\boldsymbol{D_0} &= \begin{pmatrix} x & y \\ r & u \end{pmatrix} \\
\boldsymbol{D_k} &= \begin{pmatrix} w_k & 0 \\ 0 & q_k \end{pmatrix}, \quad 1 \leq k \leq K-1 \\
\boldsymbol{D_K} &= \begin{pmatrix} -x - y - \sum_{i=1}^{K-1} w_i & 0 \\ 0 & -r - u - \sum_{i=1}^{K-1} q_i \end{pmatrix}.
\end{aligned} \tag{3.1}$$

Under this representation, the transitions where no event occurs are governed by the $\boldsymbol{D_0}$, while the transitions characterized by a batch event of size k are governed by $\boldsymbol{D_k}$. In addition, the definition of the rate matrices implies that $\boldsymbol{Q} = \sum_{k=0}^{K} \boldsymbol{D_k}$ is the infinitesimal generator of the underlying Markov process $J(t)$, with stationary probability vector $\boldsymbol{\pi} = (\pi^*, 1 - \pi^*)$, satisfying $\boldsymbol{\pi Q} = 0$ and $\boldsymbol{\pi e} = 1$, where $\boldsymbol{e}$ is a column vector of ones. The relationship between the transition probabilities matrices representation and the one based on rate matrices is

$$x = -\lambda_1, \; u = -\lambda_2, \; y = \lambda_1 p_{120}, \; r = \lambda_2 p_{210}, \; w_k = \lambda_1 p_{11k}, \; q_k = \lambda_2 p_{22k}.$$

In this paper, the characterization given by (3.1) will be the one considered from now on.

For a better understanding of the considered process, Figure 3.1 illustrates a realization of the $BMMPP_2(K)$, where the dashed line corresponds to transitions where no events occur, and the solid lines correspond to transitions where an event of size $b_i \in \{1, \ldots, K\}$ occurs.

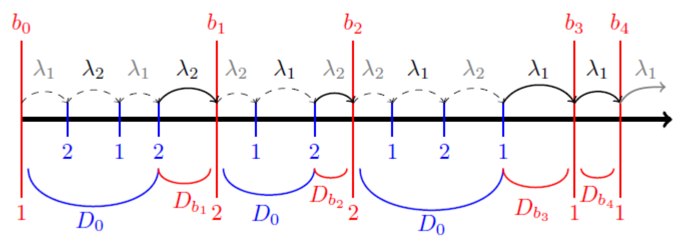


Figure 3.1: Transition diagram for the $BMMPP_2(K)$. The dashed line corresponds to transitions without events, governed by $\boldsymbol{D_0}$, and the solid lines correspond to transitions of size b_k, governed by $\boldsymbol{D_{b_k}}$.

It is important to note that if $\mathcal{B}_K = \{\boldsymbol{D_0}, \ldots, \boldsymbol{D_K}\}$ represents a *BMMPP* with maximum batch size equal to K, then $\mathcal{M} = \{\boldsymbol{G_0} = \boldsymbol{D_0},\ \boldsymbol{G_1} = \boldsymbol{D_1} + ... + \boldsymbol{D_K}\}$ defines a Markov modulated Poisson process (*MMPP*), which satisfies the same inter-event time properties as $\mathcal{B}_K$, but is not able to model events occurring in batches.

Remark 1. *Some authors define the* BMMPP *taking* $p_{iik} = p_{jjk}$ *for all* $i \neq j$*, see for example Chakravarthy (2001). In this case, the intensity matrices are expressed as* $\boldsymbol{D_0} = \boldsymbol{Q} - \boldsymbol{\Delta}(\boldsymbol{\delta})$*, where* $\boldsymbol{Q}$ *is the infinitesimal generator of the underlying Markov process* $J(t)$ *and* $\boldsymbol{\Delta}(\boldsymbol{\delta})$ *is a non-negative diagonal matrix; and* $\boldsymbol{D_k} = \boldsymbol{\Delta}(\boldsymbol{\delta})\boldsymbol{\Delta}(p_k)$*, for all* $k \geq 1$*, where* $\boldsymbol{\Delta}(p_k)$ *is a non-negative diagonal matrix with* i^{th} *diagonal entry given by* p_k*, being* $\sum_{k=1}^{K} p_k = 1$*. This is a particular case of the process introduced by Lucantoni (1993) and denoted as a* MAP *with i.i.d. batch arrivals. These processes have the advantage of being simpler than the one considered in this paper, but also have the drawback that* $corr(T, B)$ *and the first-lag autocorrelation coefficient of the inter-event times,* $\rho_B(1)$*, are both null by construction.*

(See the Appendix B for a proof of these properties). As will be seen in Section 4.2 , using the simple model in the estimation with data leads to a worse performance in modeling and its posterior use.

2.1 Performance measures regarding the inter-event times and batch sizes

A review of the performance measures concerning the inter-event times and batch sizes in a $BMMPP_2(K)$ is given next. If S_n denotes the state of the underlying Markov process at the time of the n-th event, B_n the batch size of that event and T_n the time between the $(n-1)$-th and n-th events, then the process $\{S_{n-1}, \sum_{i=1}^{n} T_i, B_n\}_{n=1}^{\infty}$, is a Markov renewal process (see for example, Chakravarthy (2010)). Furthermore, if

$$\boldsymbol{D} = \sum_{k=1}^{K} \boldsymbol{D_k},$$

then $\{S_n\}_{n=0}^{\infty}$ is a Markov chain with transition matrix

$$\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D}.$$

On the other hand, the variables T_ns are phase-type distributed with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}\}$, where $\boldsymbol{\phi}$ is the stationary probability vector associated to $\boldsymbol{P^*}$ computed as $\boldsymbol{\phi} = (\boldsymbol{\pi D e})^{-1}\boldsymbol{\pi D}$ (see Latouche and Ramaswami (1999) and Chakravarthy (2010)). In consequence, the moments of T_n in the stationary case are given by

$$\mu_r = E(T^r) = r!\boldsymbol{\phi}(-\boldsymbol{D_0})^{-r}\boldsymbol{e}, \quad \text{for } r \geq 1, \tag{3.2}$$

and the auto-correlation function of the sequence of inter-event times is

$$\rho_T(l) = \rho(T_1, T_{l+1}) = \gamma^l \frac{\mu_2 - 2\mu_1^2}{2(\mu_2 - \mu_1^2)}, \quad \text{for } l > 0. \tag{3.3}$$

In (3.3), γ is one of the two eigenvalues of the transition matrix $\boldsymbol{P^*}$ (as $\boldsymbol{P^*}$ is stochastic, then necessarily the other eigenvalue is equal to 1). According to Kang and Sung (1995), the value of γ in (3.3) is non-negative in the case of the $BMMPP_2$ and $MMPP_2$ which implies that the inter-event times are always positively correlated.

Also, from Rodríguez et al. (2016c), the mass probability function of the stationary batch size, B, is

$$P(B = k) = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D}^k\boldsymbol{e}, \quad \text{for } k = 1, ..., K,$$

from which the moments of B are obtained as

$$\beta_r = E[B^r] = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_r^*}\boldsymbol{e}, \quad \text{for } r \geq 1, \tag{3.4}$$

where $\boldsymbol{D_r^*} = \sum_{k=1}^{K} k^r \boldsymbol{D_k}$. Also, the autocorrelation function in the stationary version of the process $\rho_B(l)$ is given by

$$\rho_B(l) = \rho(B_1, B_{l+1}) = \frac{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D}]^{l-1}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e} - \beta_1^2}{\sigma_B^2},$$

where β_1 and $\sigma_B^2 = \beta_2 - \beta_1^2$ are computed from (3.4).

Using the Laplace-Stieltjes transform (LST) of the n first inter-event times and batch sizes of a stationary $BMAP_2(K)$ given in Rodríguez et al. (2016c), then $E[TB]$ is found as

$$\eta = E[TB] = \phi(-\boldsymbol{D_0})^{-2}\boldsymbol{D_1^*}\boldsymbol{e}. \tag{3.5}$$

See the Appendix A for a proof. From this, the covariance between T and B is obtained as

$$cov(T,\ B) = \phi(-\boldsymbol{D_0})^{-2}\boldsymbol{D_1^*}\boldsymbol{e} - \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e}\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}\boldsymbol{e}.$$

2.2 Performance measures regarding the counting process

Consider a stationary $BMMPP_2(K)$ represented by $\mathcal{B}_K = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \ldots, \boldsymbol{D_K}\}$ with underlying phase process $\{J(t)\}_{t\geq 0}$. Then, the counting process $N(t)$ represents the number of events that occur in $(0, t]$. For $n \in \mathbb{N}$ and $t \geq 0$, let $\boldsymbol{P}(n,t)$ denote the 2×2 matrix whose (i,j)-th element is

$$\boldsymbol{P}_{ij}(n,t) = P\left(N(t)=n,\ J(t)=j \mid N(0)=0,\ J(0)=i\right),$$

for $1 \leq i,j \leq 2$. From the previous definition it is clear that

$$p(n,t) = P\left(N(t)=n \mid N(0)=0\right) = \boldsymbol{\pi}\boldsymbol{P}(n,t)\mathbf{e}. \tag{3.6}$$

The values of the matrices $\boldsymbol{P}(n,t)$ cannot be computed in closed-form. However, their numerical computation is straightforward from the *uniformization method* addressed in Neuts and Li (1997).

If the interest is focused on counting the events of a specific size $k \in \{1,\ldots,K\}$, define $N(t,k)$ as the number of such events that have occurred up to time t. Then, it is clear that $N(t,k) \stackrel{d}{=} N^k_{\mathcal{M}}(t)$, where $N^k_{\mathcal{M}}(t)$ is the counting process of the *MMPP* given by $\mathcal{M} = \{\boldsymbol{G_0} = \boldsymbol{D_0} + \boldsymbol{D_1} + \ldots + \boldsymbol{D_{k-1}} + \boldsymbol{D_{k+1}} + \ldots + \boldsymbol{D_K},\ \boldsymbol{G_1} = \boldsymbol{D_k}\}$. Therefore, the probabilities of $N(t,k)$ can be computed as those of $N^k_{\mathcal{M}}(t)$, via expression (3.6).

Some moments concerning the counting process are as follows, see Narayana and Neuts (1992) or Eum et al. (2007). In the stationary version of the process, the mean number of events in an interval of length t (known as the *Palm function*) is

$$E\left[N(t)\right] = \lambda^\star t,$$

where $\lambda^\star = \mu_1^{-1}$ represents the events rate. The variance of that count is given by

$$\begin{aligned} V\left[N(t)\right] &= (1+2\lambda^\star)E\left[N(t)\right] - 2\boldsymbol{\pi}D\left(\mathbf{e}\boldsymbol{\pi}+\boldsymbol{Q}\right)^{-1}\boldsymbol{D}\mathbf{e}t \\ &\quad -2\boldsymbol{\pi}\boldsymbol{D}\left(\boldsymbol{I}-e^{\boldsymbol{Q}t}\right)\left(\mathbf{e}\boldsymbol{\pi}+\boldsymbol{Q}\right)^{-2}\boldsymbol{D}\mathbf{e}. \end{aligned} \tag{3.7}$$

3 Moments characterization

In this section we prove that the $BMMPP_2(K)$ is completely characterized by a set of $2(K+1)$ moments. As will be seen, the results are based on Bodrog et al. (2008), who provide a canonical representation for the MAP_2. The case where $K = 2$ shall be first addressed to later consider the generalization for an arbitrary batch size.

3.1 The $MMPP_2$, the MAP_2 and their canonical representations

As previously commented, in this paper we deal with the $BMMPP_2(K)$, which is the batch counterpart of the well-known $MMPP_2$. As described in Section 1, the $MMPP_2$ is an identifiable subclass of MAP_2, a general, non-identifiable point process that includes both renewal processes (phase type renewal processes such as the Erlang and hyperexponential renewal process) and non-renewal processes, as is the case of the $MMPP_2$. It is common in the literature to represent the MAP_2 by the rate matrices

$$\boldsymbol{G_0} = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \qquad \boldsymbol{G_1} = \begin{pmatrix} w & -x-y-w \\ v & -r-u-v \end{pmatrix}, \tag{3.8}$$

where $\{x, y, r, u, w, v\}$ are defined in a similar way to in (3.1) (see Ramírez-Cobo et al. (2010) for more details). Then, a $MMPP_2$ will be defined as (3.8) such that $w = -x - y$ and $v = 0$,

$$\boldsymbol{G_0} = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \qquad \boldsymbol{G_1} = \begin{pmatrix} -x-y & 0 \\ 0 & -r-u \end{pmatrix}. \tag{3.9}$$

Without loss of generality, we will assume from now on that $x + y \geq r + u$ (otherwise, an equivalent process is obtained by permuting the states). Note that representation (3.9) implies that events only occur at self-transitions of the underlying Markov chain, and every self-transition produces an event.

Even though the MAP_2 in (3.8) is non-identifiable, Bodrog et al. (2008) provide a canonical, unique representation; in particular, if $\gamma > 0$ (see Eq. 3.3), as is the case of the $BMMPP_2(K)$ and $MMPP_2$, then, the canonical form of (3.8) is given by

$$\boldsymbol{G_0^c} = \begin{pmatrix} -\zeta_1 & (1-a)\zeta_1 \\ 0 & -\zeta_2 \end{pmatrix}, \qquad \boldsymbol{G_1^c} = \begin{pmatrix} a\zeta_1 & 0 \\ (1-b)\zeta_2 & b\zeta_2 \end{pmatrix}, \tag{3.10}$$

for certain exponential rates ζ_1, ζ_2 and probabilities a and b. The canonical form implies that all equivalent representations of a MAP_2 as in (3.8) with associated γ satisfying $\gamma > 0$, can be written - in unique way - as in (3.10). Bodrog et al. (2008) also show that any MAP_2 as in (3.8) can be completely characterized by four moments regarding the inter-event time distribution, namely, the first, second and third moment of the inter-event time distribution, μ_1, μ_2, μ_3, and the first-lag autocorrelation coefficient of the inter-event times, $\rho_T(1)$, see Eq. (3.2) and Eq. (3.3), for their explicit expressions. Indeed, there exists a one-to-one correspondence between the parameters of the canonical form (3.10) $\{\zeta_1, \zeta_2, a, b\}$ and the moments $\{\mu_1, \mu_2, \mu_3, \rho_T(1)\}$. As will be seen in the next sections, this fact shall be the basis for proving the characterization of the $BMMPP_2(K)$ in terms of a set of moments. However, in order to find such characterization, the canonical form as in (3.10) for a $MMPP_2$ given by (3.9) needs to be found. The following result provides such a canonical representation.

Lemma 5. *Let $\mathcal{G} = \{\boldsymbol{G_0}, \boldsymbol{G_1}\}$ represent a* MMPP_2 *as in (3.9). Then representation $\mathcal{G}$ is equivalent to the canonical representation $\mathcal{G}^c = \left\{\boldsymbol{G_0^c}, \boldsymbol{G_1^c}\right\}$, where*

$$\boldsymbol{G_0^c} = \begin{pmatrix} x - y\dfrac{x-2r-u-\sqrt{(u-x)^2+4ry}}{x+2y-u+\sqrt{(u-x)^2+4ry}} & -y\dfrac{2(r+u-x-y)}{x+2y-u+\sqrt{(u-x)^2+4ry}} \\ 0 & u+y\dfrac{x-2r-u-\sqrt{(u-x)^2+4ry}}{x+2y-u+\sqrt{(u-x)^2+4ry}} \end{pmatrix},$$

$$\boldsymbol{G_1^c} = \begin{pmatrix} -x-y & 0 \\ -\dfrac{x-2r-u-\sqrt{(u-x)^2+4yr}}{2} & -r-u \end{pmatrix}.$$

Proof. The proof follows Rodríguez et al. (2016a), where a similarity transform via an invertible matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A}\mathbf{e} = \mathbf{e}$,

$$\boldsymbol{G}_{\mathbf{0}}^{\boldsymbol{c}} = \boldsymbol{A}\boldsymbol{G}_{\mathbf{0}}\boldsymbol{A}^{-1},\ \boldsymbol{G}_{\mathbf{1}}^{\boldsymbol{c}} = \boldsymbol{A}\boldsymbol{G}_{\mathbf{1}}\boldsymbol{A}^{-1} \tag{3.11}$$

is used to convert any MAP_2 as in (3.8) in its canonical form. In particular, for the $MMPP_2$, $\boldsymbol{A}^{-1}$ can be easily found as

$$\boldsymbol{A}^{-1} = \begin{pmatrix} 1 & 0 \\ -\dfrac{x-2r-u-\sqrt{(u-x)^2+4yr}}{x+2y-u+\sqrt{(u-x)^2+4yr}} & \dfrac{2(x+y-u-r)}{x+2y-u+\sqrt{(u-x)^2+4yr}} \end{pmatrix}.$$

Hence, from (3.11), the result is obtained. □

Lemma 5 shows how representation (3.10) can be obtained from (3.9). In an analogous way, the opposite transformation can be found, as the following Lemma 6 shows.

Lemma 6. *Let a* MMPP$_2$ *be represented in a canonical way by* $\mathcal{G}^c = \left\{\boldsymbol{G}_{\mathbf{0}}^{\boldsymbol{c}}, \boldsymbol{G}_{\mathbf{1}}^{\boldsymbol{c}}\right\}$ *as in (3.10). Then, its representation* $\mathcal{G} = \{\boldsymbol{G}_{\mathbf{0}}, \boldsymbol{G}_{\mathbf{1}}\}$ *as in (3.9) is given by*

$$\boldsymbol{G}_{\mathbf{0}} = \frac{1}{a\zeta_1 - b\zeta_2}\begin{pmatrix} \zeta_1\zeta_2 - a\zeta_1^2 - a\zeta_1\zeta_2 + ab\zeta_1\zeta_2 & -a^2\zeta_1^2 + a\zeta_1^2 + \zeta_2 a\zeta_1 - \zeta_2\zeta_1 \\ (\zeta_1 - b\zeta_2)(\zeta_2 - b\zeta_2) & -\zeta_1\zeta_2 + b\zeta_2^2 + b\zeta_1\zeta_2 - ab\zeta_1\zeta_2 \end{pmatrix}$$

$\boldsymbol{G}_{\mathbf{1}} = diag(G_1^c)$.

Proof. From Lemma 5 the canonical form associated to a $MMPP_2$ can be written as

$$\begin{aligned}\boldsymbol{G}_{\mathbf{0}}^{\boldsymbol{c}} &= \begin{pmatrix} -\zeta_1 & (1-a)\zeta_1 \\ 0 & -\zeta_2 \end{pmatrix} \\ &= \begin{pmatrix} x - y\dfrac{x-2r-u-\sqrt{(u-x)^2+4ry}}{x+2y-u+\sqrt{(u-x)^2+4ry}} & -y\dfrac{2(r+u-x-y)}{x+2y-u+\sqrt{(u-x)^2+4ry}} \\ 0 & u + y\dfrac{x-2r-u-\sqrt{(u-x)^2+4ry}}{x+2y-u+\sqrt{(u-x)^2+4ry}} \end{pmatrix},\end{aligned}$$

$$\boldsymbol{G_1^c} = \begin{pmatrix} a\zeta_1 & 0 \\ (1-b)\zeta_2 & b\zeta_2 \end{pmatrix}$$
$$= \begin{pmatrix} -x-y & 0 \\ -\dfrac{x-2r-u-\sqrt{(u-x)^2+4yr}}{2} & -r-u \end{pmatrix}.$$

Solving for x, y, r, u, the result is obtained. The proof that $\boldsymbol{G_0}$ and $\boldsymbol{G_1}$ are well defined can be found in Appendix C.

□

3.2 Moments characterization for the $\boldsymbol{BMMPP}_2(2)$

Consider a $BMMPP_2(2)$ represented by $\mathcal{B}_2 = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ where, according to (3.1),

$$\boldsymbol{D_0} = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \ \boldsymbol{D_1} = \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix}, \ \boldsymbol{D_2} = \begin{pmatrix} -x-y-w & 0 \\ 0 & -r-u-q \end{pmatrix}. \tag{3.12}$$

Note that $\mathcal{M} = \{\boldsymbol{G_0} = \boldsymbol{D_0},\ \boldsymbol{G_1} = \boldsymbol{D_1} + \boldsymbol{D_2}\}$ is a representation of a $MMPP_2$, and therefore, according to Lemma 1, $\mathcal{M}$ has a canonical form as in (3.10). Then, from Bodrog et al. (2008) such a canonical representation can be written in terms of $\{\mu_1, \mu_2, \mu_3, \rho_T(1)\}$, the first three inter-event time moments and the first-lag auto-correlation coefficient of the inter-event times. The next result establishes that, in order to completely characterize (3.12), two more moments involving the batch size, β_1 and η, as in (3.4) and (3.5), respectively, should be added.

Theorem 2. *Let $\mathcal{B}_2 = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ be a representation of a* $\text{BMMPP}_2(2)$ *as in (3.12). Then, $\mathcal{B}_2$ is completely characterized by the six moments $\{\mu_1, \mu_2, \mu_3, \rho_T(1), \beta_1, \eta\}$.*

Proof. Let $\mathcal{M} = \{\boldsymbol{D_0}, \boldsymbol{D_1} + \boldsymbol{D_2}\}$ be the $MMPP_2$ associated to $\mathcal{B}_2 = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$. From Lemma 6, representation $\mathcal{M}$ can be rewritten as

$$\begin{aligned}
\boldsymbol{D_0} &= \begin{pmatrix} x(\boldsymbol{\mu}, \rho_T) & y(\boldsymbol{\mu}, \rho_T) \\ r(\boldsymbol{\mu}, \rho_T) & u(\boldsymbol{\mu}, \rho_T) \end{pmatrix} \\
\boldsymbol{D_1} + \boldsymbol{D_2} &= \begin{pmatrix} -x(\boldsymbol{\mu}, \rho_T) - y(\boldsymbol{\mu}, \rho_T) & 0 \\ 0 & -r(\boldsymbol{\mu}, \rho_T) - u(\boldsymbol{\mu}, \rho_T) \end{pmatrix},
\end{aligned}$$

where $\boldsymbol{\mu} = \{\mu_1, \mu_2, \mu_3\}$. Hence

$$\begin{aligned}
\boldsymbol{D_0} &= \begin{pmatrix} x(\boldsymbol{\mu}, \rho_T) & y(\boldsymbol{\mu}, \rho_T) \\ r(\boldsymbol{\mu}, \rho_T) & u(\boldsymbol{\mu}, \rho_T) \end{pmatrix} \\
\boldsymbol{D_1} &= \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix} \\
\boldsymbol{D_2} &= \begin{pmatrix} -x(\boldsymbol{\mu}, \rho_T) - y(\boldsymbol{\mu}, \rho_T) - w & 0 \\ 0 & -r(\boldsymbol{\mu}, \rho_T) - u(\boldsymbol{\mu}, \rho_T) - q \end{pmatrix}.
\end{aligned}$$

The quantities β_1 and η defined in (3.4) and (3.5) respectively, can be written in the case of the $BMMPP_2(K)$ as

$$\beta_1 = \frac{2(rx + 2ry + yu) + rw + yq}{(rx + 2ry + yu)} \tag{3.13}$$

and

$$\eta = \frac{rw(y - u) + qy(r - x) + (ry - xu)(2r + 2y)}{(rx + 2ry + yu)(xu - ry)}. \tag{3.14}$$

From (3.13)

$$rw = (\beta_1 - 2)(rx + ry + ry + yu) - yq \tag{3.15}$$

and from substituting (3.15) in (3.14),

$$\begin{aligned}
\eta &= \frac{[(\beta_1-2)(rx+2ry+yu)-yq](y-u)}{(rx+2ry+yu)(xu-ry)} \\
&\quad + \frac{qy(r-x)+(ry-xu)(2r+2y)}{(rx+2ry+yu)(xu-ry)} \\
&= \frac{(\beta_1-2)(rx+2ry+yu)(y-u)}{(rx+2ry+yu)(xu-ry)} \\
&\quad + \frac{qy(r+u-x-y)+(ry-xu)(2r+2y)}{(rx+2ry+yu)(xu-ry)}.
\end{aligned}$$

Hence

$$\begin{aligned}
q &= \frac{\eta(rx+2ry+yu)(xu-ry)-(\beta_1-2)(rx+2ry+yu)(y-u)}{y(r+u-x-y)} \\
&\quad - \frac{(ry-xu)(2r+2y)}{y(r+u-x-y)} \\
&= \frac{(rx+2ry+yu)[(xu-ry)\eta-(y-u)(\beta_1-2)]}{y(r+u-x-y)} \\
&\quad - \frac{(ry-xu)(2r+2y)}{y(r+u-x-y)} \qquad (3.16)
\end{aligned}$$

and from substituting (3.16) in (3.15), w is finally found as

$$w = \frac{(rx+2ry+yu)[(\beta_1-2)(r-x)-(xu-ry)\eta]+(ry-xu)(2r+2y)}{r(r+u-x-y)}$$

Since the parameters defining $\mathcal{B}_2$ are written in terms of the moments $\{\mu_1, \mu_2, \mu_3, \rho_T(1), \beta_1, \eta\}$, the proof is completed. □

3.3 The case $K \geq 3$

In this section the characterization in terms of moments is extended from the case $K = 2$ to the case with an arbitrary maximum batch size K. The key for such generalization is the fact that given a $BMMPP_2(K)$ represented by $\mathcal{B}_K = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K}\}$, then K different $BMMPP_2(2)$s can be obtained as

$$\mathcal{B}_2^{(i)} = \{\boldsymbol{D_0}, \boldsymbol{D_i}, \sum_{k \neq i} \boldsymbol{D_k}\}, \quad i = 1, ..., K. \tag{3.17}$$

Theorem 3. *Let $\mathcal{B}_K = \{\boldsymbol{D_0}, \boldsymbol{D_1}, ..., \boldsymbol{D_K}\}$ be the representation of a* BMMPP$_2(K)$*. Then, $\mathcal{B}_K$ is characterized by the set of $(2K+2)$ moments*

$$\left\{\mu_1, \mu_2, \mu_3, \rho_T(1), \beta_1^{(1)}, \eta^{(1)}, \ldots, \beta_1^{(K-1)}, \eta^{(K-1)}\right\}, \tag{3.18}$$

where $\beta_1^{(i)}$ and $\eta^{(i)}$ are the moments defined according to (3.4) and (3.5) of $\mathcal{B}_2^{(i)}$, for $i = 1, \ldots, (K-1)$, that is the BMMPP$_2(2)$ *as in (3.17).*

Proof. The proof is straightforward by applying Theorem 2 to each one of the $BMMPP_2(2)$s defined by $\mathcal{B}_2^{(i)}$, as in (3.17), for $i = 1, \ldots, (K-1)$. □

4 Inference for the $\boldsymbol{BMMPP}_2(K)$

In this section, an approach for estimating the parameters of a $BMMPP_2(K)$ given observed inter-event times, $\boldsymbol{t} = (t_1, t_2, ..., t_n)$ and batch sizes, $\boldsymbol{b} = (b_1, b_2, ..., b_n)$, is proposed. This implies that some components of the process such as the complete sequence of transition times and the sequence of visited states in the underlying Markov process are not observed, which corresponds with what usually occurs in practice.

Section 4.1 presents in detail the novel fitting algorithm, where the rate matrices $\boldsymbol{D_0}, \ldots, \boldsymbol{D_K}$ are sequentially estimated via $(K+1)$ optimization problems solved by standard optimization routines. Then, Section 4.2 illustrates the performance of the method on simulated data sets and Section 4.3 compares the novel approach with an EM-based strategy proposed in the literature. Finally, Section 4.4 addresses the modeling of the well-known Bellcore Aug89 data set, where in addition, a performance analysis related to the $BMMPP/M/1$ queueing system is considered.

4.1 The fitting algorithm

Theorem 3 shows that any $BMMPP_2(K)$, with rate matrix representation as in (3.1), is characterized by the set of $2(K+1)$ moments given by (3.18).

Specifically, such moments are given by μ_1, μ_2, μ_3, and $\rho_T(1)$ (concerning the inter-event time distribution), $\beta_1^{(1)}, \ldots, \beta_1^{(K-1)}$ (related to the batch size distribution), and $\eta^{(1)}, \ldots, \eta^{(K-1)}$ (joint moments concerning times and sizes). Carrizosa and Ramírez-Cobo (2014) derive a moments matching method for estimating the parameters of a MAP_2 as in (3.8), given a sequence of inter-event times $\boldsymbol{t} = (t_1, t_2, ..., t_n)$. A modified version of the approach in Carrizosa and Ramírez-Cobo (2014) shall constitute the first step in our sequential fitting algorithm aimed to estimate matrix $\boldsymbol{D_0}$. Any $BMMPP_2(K)$ given by $\mathcal{B}_2 = \{\boldsymbol{D_0}, \boldsymbol{D_1}, \ldots, \boldsymbol{D_K}\}$ defines a $MMPP_2$ represented by $\mathcal{M} = \{\boldsymbol{G_0} = \boldsymbol{D_0},\ \boldsymbol{G_1} = \sum_{k=1}^{K} \boldsymbol{D_k}\}$, where $\boldsymbol{G_0}$ and $\boldsymbol{G_1}$ are as in (3.9); therefore, $\boldsymbol{G_0}$ ($\boldsymbol{D_0}$) can be estimated by the solution of the following moments matching optimization problem ($P0$):

$$
(P0)\begin{cases}
\min\limits_{x,y,r,u} & \delta_{0,\tau}(x,y,r,u) \\
\text{s.t.} & x, u \leq 0, \\
 & y, r \geq 0, \\
 & x + y \leq 0, \\
 & r + u \leq 0,
\end{cases}
$$

where the objective function is

$$
\begin{aligned}
\delta_{0,\tau}(x,y,r,u) &= \{\rho_T(1) - \bar{\rho}_T(1)\}^2 + \\
&+ \tau\left\{\left(\frac{\mu_1 - \bar{\mu}_1}{\bar{\mu}_1}\right)^2 + \left(\frac{\mu_2 - \bar{\mu}_2}{\bar{\mu}_2}\right)^2 + \left(\frac{\mu_3 - \bar{\mu}_3}{\bar{\mu}_3}\right)^2\right\},
\end{aligned}
$$

for a value of τ to be tuned in practice, and where $\bar{\mu}_i$, for $i = 1, 2, 3$ and $\bar{\rho}_T(1)$ denote the empirical moments (computed from the sample $\boldsymbol{t}$). Note that in the previous objective function, $\rho_T(1) = \rho_T(1)(x, y, r, u)$, and $\mu_i = \mu_i(x, y, r, u)$, for $i = 1, 2, 3$.

Once $\hat{\boldsymbol{D}}_{\boldsymbol{0}}$ is obtained as the solution of (P0), then, in order to estimate $\boldsymbol{D_1}$ (or equivalently w_1, q_1), consider (3.17) for $i = 1$, that is, the

$BMMPP_2(2)$ represented by $\mathcal{B}_2^{(1)} = \{\boldsymbol{D_0},\ \boldsymbol{D_1},\ \boldsymbol{D_2} + \ldots + \boldsymbol{D_K}\}$ and the optimization problem

$$(P1) \begin{cases} \min\limits_{w_1, q_1} & \delta_{1,\tau}(\hat{x}, \hat{y}, \hat{u}, \hat{v}, w_1, q_1) \\ \text{s.t.} & 0 \leq w_1 \leq -\hat{x} - \hat{y}, \\ & 0 \leq q_1 \leq -\hat{r} - \hat{u}, \end{cases}$$

where, according to (3.1), $\hat{x}, \hat{y}, \hat{r}, \hat{u}$ are the elements of $\boldsymbol{\hat{D}_0}$ and

$$\delta_{1,\tau}(x, y, u, v, w_1, q_1) = \tau \left\{ \left(\frac{\beta_1^{(1)} - \bar{\beta}_1^{(1)}}{\bar{\beta}_1^{(1)}} \right)^2 + \left(\frac{\eta^{(1)} - \bar{\eta}^{(1)}}{\bar{\eta}^{(1)}} \right)^2 \right\}. \quad (3.19)$$

In the previous objective function (3.19), $\beta_1^{(1)} = \beta_1^{(1)}(x, y, u, r, w_1, q_1)$ and similarly, $\eta^{(1)} = \eta^{(1)}(x, y, u, r, w_1, q_1)$. It is crucial to remark that, in order to compute the empirical moments $\bar{\beta}_1^{(1)}$ and $\bar{\eta}^{(1)}$, all batch sizes in $\boldsymbol{b}$ larger than 2 are considered as equal to 2. Once $\hat{w}_1$ and $\hat{q}_1$ are obtained as the solutions of $(P1)$, the approach will be repeated for estimating $\boldsymbol{D_2}$ (using the representation of $\mathcal{B}_2^{(2)}$), $\boldsymbol{D_3}$,..., and finally $\boldsymbol{D_K}$. The algorithm is summarized in Table 3.1.

It is important to comment that the optimization problems (Pk) in Table 3.1 for $k = 0, \ldots, K-1$ are straightforward problems in two variables each, solved using standard optimization routines (`fmincon` in MATLAB©), where a multistart with 100 randomly chosen starting points was executed.

4.2 A simulational study

The aim of this section is twofold: on one hand, the behavior of the sequential algorithm described in Section 4 is illustrated on the basis of two simulated data sets and, on the other hand, a sensitivity analysis concerning the tuning parameter τ is undertaken. Each simulated data set consists of a sequence of inter-event times $\boldsymbol{t} = (t_1, t_2, ..., t_n)$ and a sequence of batch sizes $\boldsymbol{b} = (b_1, b_2, ..., b_n)$. The first data set was simulated from the $BMMPP_2(2)$

1. `Obtain` $(\hat{x}, \hat{y}, \hat{r}, \hat{u})$ `(equivalently,` $\hat{\boldsymbol{D}}_{\mathbf{0}}$`) as the solution of (P0).`

2. `For` $k = 1, \ldots, K-1$ `repeat:`

   (a) `Compute the empirical moments` $\bar{\beta}_1^{(k)}$ `and` $\bar{\eta}^{(k)}$ `from` $\boldsymbol{t}$ `and the sample of baches` $\boldsymbol{b}^\star = (b_1^\star, \ldots, b_n^\star)$`, where for` $j = 1, \ldots, n$`,` $b_j^\star = 1$ `if` $b_j = k$`, or` $b_j^\star = 2$`, otherwise.`

   (b) `From` $\hat{\boldsymbol{D}}_{\mathbf{0}}, \ldots, \hat{\boldsymbol{D}}_{\boldsymbol{k-1}}$ `and the moments` $\hat{\beta}_1^{(k)}$`,` $\hat{\eta}^{(k)}$`, obtain` $\hat{w}_k, \hat{q}_k$ `(`$\hat{\boldsymbol{D}}_{\boldsymbol{k}}$`) as the solutions of`

$$(Pk) \left\{ \begin{array}{ll} \displaystyle\min_{w_k, q_k} & \delta_{k,\tau}(\hat{x}, \hat{y}, \hat{u}, \hat{v}, \hat{w}_1, \hat{q}_1, \ldots, \hat{w}_{k-1}, \hat{q}_{k-1}, w_k, q_k) \\ \texttt{s.t.} & 0 \leq w_k \leq -\left(\hat{x} + \hat{y} + \hat{w}_1 + \ldots + \hat{w}_{k-1}\right), \\ & 0 \leq q_k \leq -\left(\hat{r} + \hat{u} + \hat{q}_1 + \ldots + \hat{q}_{k-1}\right), \end{array} \right.$$

   `where`

$$\delta_{k,\tau}(w_k, q_k) = \tau \left\{ \left(\frac{\beta_1^{(k)} - \bar{\beta}_1^{(k)}}{\bar{\beta}_1^{(k)}} \right)^2 + \left(\frac{\eta^{(k)} - \bar{\eta}^{(k)}}{\bar{\eta}^{(k)}} \right)^2 \right\}.$$

Table 3.1: Sequential algorithm for estimating the $BMMPP_2(K)$ parameters

represented by the rate matrices $\{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}\}$ shown in the second column of Table 3.2; the second one was generated from the $BMMPP_2(4)$ characterized by $\{\boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{D_2}, \boldsymbol{D_3}, \boldsymbol{D_4}\}$ as in the second column of Table 3.3. An important remark concerning the samples sizes n needs to be made at this point. The estimation approach proposed in Section 4 uses as input arguments a set of empirical moments concerning both the inter-event times and batch sizes. Since the process is known to be identifiable (Yera et al., 2019) then, the closer the empirical moments are to the theoretical moments, the more accurate the estimated parameters will be. Therefore, the issue of the sample size is critical in this context. In this paper, we adopt the approach as in Ramírez-Cobo et al. (2017), where the coefficient of variation of the inter-event times is taken into account. Specifically, if the coefficient of variation is high, then the sequence of inter-event times will present more variability and therefore, the approximation of the empirical moments to the theoretical ones may be poor. Under the two generator processes considered, the coefficients of variation are equal to 1.02 and 2.048, respectively. Similarly as in Ramírez-Cobo et al. (2017), we fix a lower value of the sample size ($n = 300$) for the first case than for the second case ($n = 1000$).

The results obtained when the novel estimation approach is used to fit the traces are shown in Tables 3.2 and 3.3. Tables 3.2 is related to the simulated sample from the $BMMPP_2(2)$ while the Tables 3.3 concerns the second simulated sample from the $BMMPP_2(4)$. The second column in the Tables 3.2 and Tables 3.3 show the generator process and the characterizing theoretical moments according to Sections 3.2 and 3.3. The third column in these Tables show the empirical moments from the simulated traces. The rest of columns show the estimated rate matrices and estimated characterizing moments under the new approach for an assortment of values of the tuning parameter τ ($\tau \in \{0.001,\ 0.01,\ 0.1,\ 1,\ 10,\ 100\}$). Finally, the last row in both Tables shows the running time (measured in seconds) employed for the novel method in an Intel Core i5 of dual-core 2.6 GHz processor with 4Gb of memory ram (for a prototype code written in MATLAB©).

Some comments arise from the results presented in Tables 3.2 and 3.3. First, from the third column it can be concluded that the selected sample sizes ($n = 300$ and $n = 1000$) are good enough to guarantee an accurate approximation of the empirical moments to the theoretical ones. Second, the value of τ does not seem to affect the estimation significantly: both the rate matrices and estimated moments are close to the real ones in all cases. However, the value of τ seems to have an impact on the computational time: the lower τ is, the faster the method turns out to be. For this reason, the smallest tested value ($\tau = 0.001$) will be considered from now on in the rest of the experiments. Finally, it is important to note that the sample size does not affect the running times, a fact which in any case was expected since the input arguments of the algorithm are empirical moments (and not the original traces).

The choice of processes in Tables 3.2 and 3.3 is also related to Remark

1. As can be observed, the process analyzed in Table 3.2 presents autocorrelation between batches, autocorrelation between the inter-arrivals times and correlation between T and B close to zero; while in the second process, studied in Table 3.3, they are significantly different from zero. The difference between fitting a *BMMPP* as defined in this work and the *MAP* with i.i.d. batch arrivals is more relevant in the second model than in the first as is illustrated in Table 3.4. It can be seen that the quality of the performance in the adjustment is better using the methodology developed in this paper, while the computational times are very similar.

			τ					
	Generator Process	*Empirical*	0.001	0.01	0.1	1	10	100
$\boldsymbol{D_0}$	$\begin{pmatrix} -5 & 2 \\ 5 & 10 \end{pmatrix}$	-	$\begin{pmatrix} -5.95 & 3.07 \\ 6.59 & -11.64 \end{pmatrix}$	$\begin{pmatrix} -5.95 & 3.08 \\ 6.46 & -11.50 \end{pmatrix}$	$\begin{pmatrix} -5.92 & 3.04 \\ 6.52 & -11.57 \end{pmatrix}$	$\begin{pmatrix} -6.22 & 3.40 \\ 6.30 & -11.25 \end{pmatrix}$	$\begin{pmatrix} -5.86 & 2.98 \\ 6.60 & -11.68 \end{pmatrix}$	$\begin{pmatrix} -5.78 & 2.87 \\ 7.64 & -12.95 \end{pmatrix}$
$\boldsymbol{D_1}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$	-	$\begin{pmatrix} 0.91 & 0 \\ 0 & 2.08 \end{pmatrix}$	$\begin{pmatrix} 0.91 & 0 \\ 0 & 2.07 \end{pmatrix}$	$\begin{pmatrix} 0.91 & 0 \\ 0 & 2.08 \end{pmatrix}$	$\begin{pmatrix} 0.89 & 0 \\ 0 & 2.02 \end{pmatrix}$	$\begin{pmatrix} 0.92 & 0 \\ 0 & 2.09 \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0 \\ 0 & 2.23 \end{pmatrix}$
$\boldsymbol{D_2}$	$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$	-	$\begin{pmatrix} 1.96 & 0 \\ 0 & 2.98 \end{pmatrix}$	$\begin{pmatrix} 1.96 & 0 \\ 0 & 2.96 \end{pmatrix}$	$\begin{pmatrix} 1.96 & 0 \\ 0 & 2.97 \end{pmatrix}$	$\begin{pmatrix} 1.94 & 0 \\ 0 & 2.92 \end{pmatrix}$	$\begin{pmatrix} 1.97 & 0 \\ 0 & 2.99 \end{pmatrix}$	$\begin{pmatrix} 1.98 & 0 \\ 0 & 3.08 \end{pmatrix}$
μ_1	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
μ_2	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
μ_3	0.138	0.138	0.138	0.138	0.138	0.138	0.138	0.138
$\rho_T(1)$	7.35×10^{-3}	5.99×10^{-3}	5.99×10^{-3}	6.01×10^{-3}	6.03×10^{-3}	5.81×10^{-3}	6.07×10^{-3}	5.99×10^{-3}
β_1	1.64	1.64	1.64	1.64	1.64	1.64	1.64	1.64
η_1	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
$\rho_B(1)$	2.24×10^{-3}	1.73×10^{-3}	1.02×10^{-3}	8.82×10^{-4}	8.28×10^{-4}	8.30×10^{-4}	8.39×10^{-4}	8.39×10^{-4}
$corr(T,B)$	5.83×10^{-3}	7.53×10^{-3}	7.02×10^{-3}	7.51×10^{-4}	7.50×10^{-3}	7.53×10^{-3}	7.53×10^{-3}	7.53×10^{-3}
running time	-	-	21.59	35.24	51.40	108.64	180.69	183.96

Table 3.2: Performance of the novel sequential estimation method for a simulated trace from a $BMMPP_2(2)$ for an assortment of τ values.

			τ					
	Generator Process	*Empirical*	0.001	0.01	0.1	1	10	100
$\boldsymbol{D_0}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.91 & -14.20 \end{pmatrix}$	-	$\begin{pmatrix} -0.58 & 0.09 \\ 1.85 & -13.86 \end{pmatrix}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.84 & -13.82 \end{pmatrix}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.84 & -13.79 \end{pmatrix}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.84 & -13.78 \end{pmatrix}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.84 & -13.79 \end{pmatrix}$	$\begin{pmatrix} -0.58 & 0.09 \\ 1.83 & -13.78 \end{pmatrix}$
$\boldsymbol{D_1}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.47 \end{pmatrix}$	-	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.17 \end{pmatrix}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.14 \end{pmatrix}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.14 \end{pmatrix}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.14 \end{pmatrix}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.14 \end{pmatrix}$	$\begin{pmatrix} 0.08 & 0 \\ 0 & 11.14 \end{pmatrix}$
$\boldsymbol{D_2}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.10 \end{pmatrix}$	-	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.08 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.07 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.07 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.07 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.07 \end{pmatrix}$	$\begin{pmatrix} 0.15 & 0 \\ 0 & 0.07 \end{pmatrix}$
$\boldsymbol{D_3}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.60 \end{pmatrix}$	-	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0 \\ 0 & 0.64 \end{pmatrix}$
$\boldsymbol{D_4}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.12 \end{pmatrix}$	-	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.10 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.11 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.11 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.11 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.11 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0 \\ 0 & 0.11 \end{pmatrix}$
μ_1	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97
μ_2	3.34	3.26	3.26	3.26	3.26	3.26	3.26	3.26
μ_3	17.65	17.08	17.08	17.08	17.08	17.08	17.08	17.08
$\rho_T(1)$	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
β_1	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42
β_2	1.86	1.86	1.86	1.86	1.86	1.86	1.86	1.86
β_3	1.73	1.74	1.74	1.74	1.74	1.74	1.74	1.74
η_1	1.67	1.65	1.65	1.65	1.65	1.65	1.65	1.65
η_1	1.71	1.68	1.68	1.68	1.68	1.68	1.68	1.68
η_3	1.53	1.54	1.52	1.52	1.52	1.52	1.52	1.52
$\rho_B(1)$	0.36	0.36	0.28	0.27	0.27	0.27	0.27	0.27
$Corr(T,B)$	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
running time	-	-	37.04	65.88	76.57	159.58	171.36	179.02

Table 3.3: Performance of the novel sequential estimation method for a simulated trace from a $BMMPP_2(4)$ for an assortment of τ values.

	Emp	Est BMMPP	Est MMPP with i.i.d. batches	Emp	Est BMMPP	Est MMPP with i.i.d. batches
μ_1	0.2801	0.2797	0.2798	0.9657	0.9657	0.9656
μ_2	0.1602	0.1602	0.1604	3.2627	3.2628	3.2635
μ_3	0.1382	0.1382	1.381	17.0847	17.0842	17.0827
$\rho_T(1)$	5.990×10^{-3}	5.989×10^{-3}	5.989×10^{-3}	0.2241	0.2241	0.2241
$\beta_1^{(1)}$	1.6397	1.6397	1.6423	1.4182	1.4182	1.5358
$\eta^{(1)}$	0.4603	0.4603	0.4595	1.6472	1.6472	1.4830
$\beta_1^{(2)}$	-	-	-	1.8563	1.8558	1.7962
$\eta^{(2)}$	-	-	-	1.6833	1.6835	1.7345
$\beta_1^{(3)}$	-	-	-	1.7389	1.7390	1.6681
$\eta^{(3)}$	-	-	-	1.5148	1.5148	1.6107
CV	1.0208	1.0218	1.0242	1.5806	1.5807	1.5812
Skewness	5.9111	5.9004	5.8675	4.8034	4.8030	4.7997
Kurtosis	23.7610	27.7867	23.7164	22.0435	21.9988	21.9887
β_1	1.6397	1.6397	1.6423	1.7064	1.7062	1.8679
β_2	2.9190	2.9190	2.9270	3.7224	3.7217	4.2679
$Corr(T,B)$	7.536×10^{-3}	8.734×10^{-3}	0	0.3268	0.3270	0
$\rho_B(1)$	2.248×10^{-3}	1.025×10^{-3}	0	0.3633	0.2665	0
$\rho_B(2)$	3.577×10^{-4}	2.907×10^{-4}	0	0.2635	0.1991	0
$\rho_B(3)$	7.479×10^{-4}	8.247×10^{-5}	0	0.2060	0.1488	0
$\rho_T(2)$	6.426×10^{-4}	1.699×10^{-3}	1.537×10^{-3}	0.1652	0.1675	0.1674
$\rho_T(3)$	1.755×10^{-3}	5.819×10^{-4}	3.943×10^{-4}	0.1211	0.1251	0.1250
$P(B=1)$	0.3603	0.3603	0.3577	0.5818	0.5818	0.4642
$P(B=2)$	0.6397	0.6397	0.6423	0.1437	0.1437	0.2038
$P(B=3)$	-	-	-	0.2611	0.2611	0.3319
$P(B=4)$	-	-	-	0.0135	0.0135	6.932×10^{-5}
running time	-	21.59	19.55	-	37.04	33.13

Table 3.4: Comparisson between the estimated descriptors via the *BMMPP*) and the *MMPP* with i.i.d. batches.

4.3 Comparison with the EM algorithm and estimation of the $BMMPP_2(K)/M/1$ queue

This section serves two purposes. First, as commented in Section 1, some authors have considered inference for the general *BMAP*s, such as Breuer

(2002) and Klemm et al. (2003) who adapt the EM algorithm for the $BMAP$. Therefore, one of the aims of this section is to compare the performance of the novel sequential fitting methods with that of the EM algorithm as implemented in Breuer (2002). Second, one of the main applications of $BMAP$ processes are related to queueing theory, see for example Lucantoni et al. (1990); Ramaswami (1990); Lucantoni (1991, 1993); Lucantoni et al. (1994), who explore theoretical properties of the $BMAP/G/1$ queueing system. In this section, we consider estimation for the $BMMPP_2(K)/M/1$ queueing system where the $BMMPP_2(K)$ is the arrival process in a single-server, first in first out queueing system with independent, Markovian service times. In particular, the inference approach described in Section 4 will be combined with techniques from the queueing literature in order to estimate the stationary queue length distribution at departures.

In Breuer (2002) and Klemm et al. (2003), the EM algorithm is considered and adapted for the $BMAP$. In order to compare the performance of the novel method with that of the EM algorithm, we consider the first simulated trace from Section 4.2, with generator process and theoretical moments as in the second column of Table 3.5. To explore in depth the performance of the EM algorithm, two different starting points are considered; the first quite close to the true solution (fourth column of Table 3.5) and a second point that is far away from the true solution (seventh column of Table 3.5). From the results in the table, some conclusions can be obtained. First, the estimated rate matrices provided by the EM algorithms seem more dependent on the starting solution than those under the novel approach; while the solutions obtained with the moments matching method are similar under the two choices of the initial values, the solutions given by the EM differ among themselves, with the first one being more accurate than the second one. Concerning the estimation of the empirical moments, both methods provide similar values, all close to the empirical ones. Something similar occurs with respect to the log-likelihood values given the estimated parameters (second-to-last row of Table 3.5). Finally, concerning the running times, the EM algorithm turns out to be notably slower than the novel method, especially when the starting solution is not close to the true one.

			Close starting point			Distant starting point		
	Generator Process	*Empirical*	*Starting solution*	*EM*	*Sequential approach*	*Starting solution*	*EM*	*Sequential approach*
$\boldsymbol{D_0}$	$\begin{pmatrix} -5 & 2 \\ 5 & -10 \end{pmatrix}$	-	$\begin{pmatrix} -10 & 3 \\ 5 & -15 \end{pmatrix}$	$\begin{pmatrix} -5.05 & 2.22 \\ 6.80 & -12.07 \end{pmatrix}$	$\begin{pmatrix} -4.15 & 0.89 \\ 6.49 & -12.24 \end{pmatrix}$	$\begin{pmatrix} -25 & 10 \\ 13 & -27 \end{pmatrix}$	$\begin{pmatrix} -11.51 & 8.42 \\ 16.55 & -20.61 \end{pmatrix}$	$\begin{pmatrix} -4.21 & 0.98 \\ 4.54 & -9.60 \end{pmatrix}$
$\boldsymbol{D_1}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$	-	$\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$	$\begin{pmatrix} 0.73 & 0 \\ 0 & 3.02 \end{pmatrix}$	$\begin{pmatrix} 1.15 & 0 \\ 0 & 2.77 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 0.47 & 0 \\ 0 & 2.85 \end{pmatrix}$	$\begin{pmatrix} 1.27 & 0 \\ 0 & 1.63 \end{pmatrix}$
$\boldsymbol{D_2}$	$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$	-	$\begin{pmatrix} 3 & 0 \\ 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 2.09 & 0 \\ 0 & 2.25 \end{pmatrix}$	$\begin{pmatrix} 2.10 & 0 \\ 0 & 2.98 \end{pmatrix}$	$\begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}$	$\begin{pmatrix} 2.62 & 0 \\ 0 & 1.21 \end{pmatrix}$	$\begin{pmatrix} 1.96 & 0 \\ 0 & 3.43 \end{pmatrix}$
μ_1	0.28	0.28	-	0.29	0.28	-	0.29	0.28
μ_2	0.16	0.16	-	0.17	0.16	-	0.17	0.16
μ_3	0.138	0.137	-	0.158	0.138	-	0.148	0.138
$\rho_T(1)$	7.35×10^{-3}	6.01×10^{-3}	-	2.92×10^{-2}	5.99×10^{-3}	-	1.16×10^{-1}	6.01×10^{-3}
β_1	1.64	1.62	-	1.67	1.64	-	1.82	1.62
η_1	0.46	0.46	-	0.49	0.46	-	0.53	0.46
l	-	-	-270.66	-114.37	-114.63	-414.25	-114.10	-114.70
running time	-	-	-	52.21	0.23	-	98.72	0.59

Table 3.5: Comparison between the EM algorithm and the novel sequential approach under two different starting solutions.

Consider next the $BMMPP_2(K)/M/1$ queueing system and denote by $\mu^\star < \infty$ the expected value of the service time. Then, the traffic intensity of this system is given by

$$\rho = \lambda^\star \mu^\star,$$

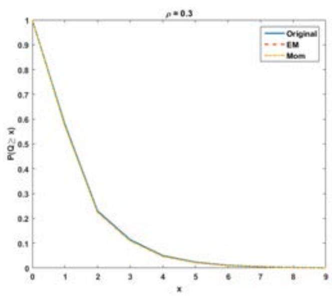
where $\lambda^\star$ is the stationary arrival rate (inverse of the expected inter-event time), defined as

$$\lambda^\star = 1/\mu_1,$$

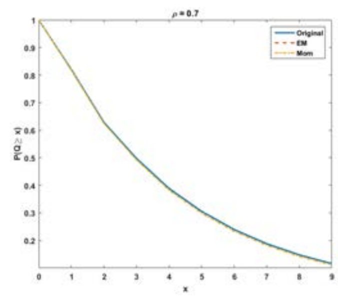
where μ_1 is defined as in (3.2). Now define $Z(t)$ to be the number of customers in the system (including in service, if any) at time t and let τ_k be the epoch of the k-th departure from the queue, with $\tau_0 = 0$. If the system is stable ($\rho < 1$), then for $i \geq 1$

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

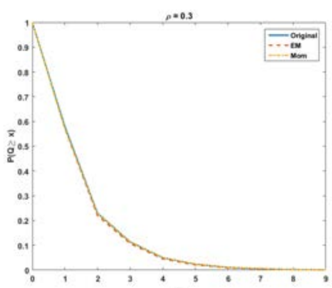
represents the stationary probability that the queue length is equal to i when a departure occurs. Closed-form expressions for the generating function of the queue length distributions can be found in Lucantoni (1993). Assume that the simulated trace of inter-event times used in Table 3.5 represents the inter-arrival times in a $BMMPP_2(2)/M/1$ queue. Then, given the point estimates of the $BMMPP_2(2)$ from the table, the numerical routines described in Lucantoni (1993), as well as in Abate and Whitt (1995) can be implemented to invert the generating function of the queue length distribution. Figure 3.2 depicts the estimated tail distributions of the queue length at departures for two different service times (that is, for two different traffic intensities, $\rho = 0.3,\ 0.7$) and for both solutions (from the sequential fitting approach and EM algorithm) and under the two possible choices of starting points considered in Table 3.5. In the figure, the solid line represents the true distribution, the dashed line is the estimated function using the EM solution and finally, the dotted line depicts the estimated tail distribution under the solution obtained by the sequential fitting method. From the figure some comments can be made. First, as expected, larger values for the tail distribution are obtained in the case of $\rho = 0.7$, a consequence of the higher degree of saturation of the system. Second, an additional expected fact is that the estimated tail distributions using a close starting point are slightly more accurate than those obtained under the distant starting points. Finally, in the case of the distant starting point with $\rho = 0.7$, the sequential fitting approach leads to a slightly more precise solution than the EM algorithm.



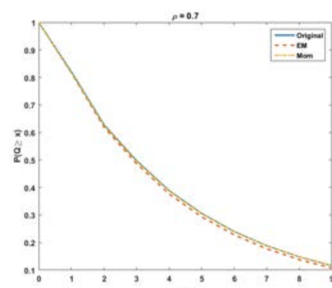
(a) Close starting point, $\rho = 0.3$



(b) Close starting point, $\rho = 0.7$



(c) Distant starting point, $\rho = 0.3$



(d) Distant starting point, $\rho = 0.7$

Figure 3.2: Estimated tail distributions of the queue length at departures in a $BMMPP_2(2)/M/1$ queue.

4.4 Numerical illustration on a teletraffic real data set

In this section we illustrate the performance of the novel approach for fitting a well-referenced database, namely the Bellcore Aug89 data set, which has been considered in a number of papers concerning teletraffic modeling, see Horváth et al. (2005),Ramírez-Cobo et al. (2008); Ramírez-Cobo et al.

(2010); Li et al. (2010); Kriege and Buchholz (2011); Okamura et al. (2011); Rodríguez et al. (2015); Casale et al. (2016).

Data description

The data set BC-pAug89, available at the web site

`http://ita.ee.lbl.gov/html/contrib/BC.html`

consists of one million of packet arrivals seen on Ethernet at the Bellcore Morristown Research and Engineering facility. The trace began at 11:25 on August 29, 1989, and ran for about 3142.82 seconds until the arrival of one million packets (of different size each). The times are originally expressed to 6 places after the decimal point (milisecond resolution), which implies that packets arrive in isolated way. However, if instead of observing the process every 10^{-6} seconds, it is observed every 10^{-3} seconds, then this form of aggregation leads to packets arriving in batches, with batches sizes varying from 1 to 4, as shown by the left panel of Figure 3.3. This structure of the data set will be called from now on Data set in format I. On the other hand, the original data set can be viewed from a different perspective if the size of the packets is taken into account. Due to the Ethernet protocol the size of the packets takes 866 different values ranging from 64 to 1518 bytes. Therefore, packets can be divided into *small* packets, when the size is lower than 100 bits, and *large*, otherwise, see the right panel of Figure 3.3. This new format, proposed in an analogous way in Klemm et al. (2003), will be called henceforth Data set in format II, where the batch size equal to 1 will refer to *small* sizes, and a batch size of 2 will be used to refer to the *large* sizes.

Consider first Data set in format I. There are strong reasons to not assume a Poisson process for the Data set in format I . First, the average, median, variation coefficient, minimum and maximum value of the inter-arrivals times are 0.0036, 0.0020, 1.6553, 1×10^{-3} and 0.3420 seconds, respectively, which suggests a right-skewed distribution with a tail longer than that of an exponential distribution. Indeed, Figure 3.4 shows the empirical quantiles

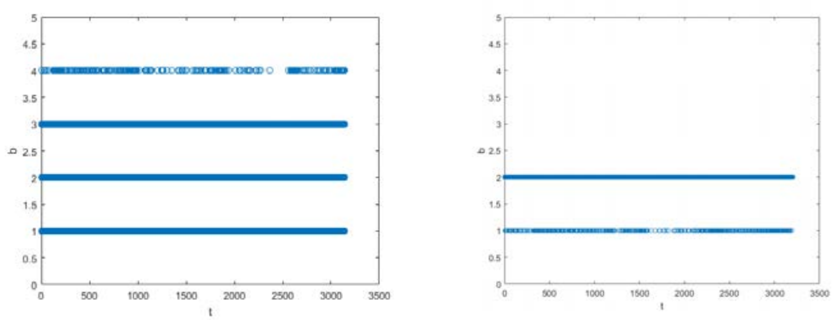


Figure 3.3: Left panel: packets arriving in batches of sizes $1, \ldots, 4$, observed in intervals of length 10^{-3} seconds (Data set in format I). Right panel: packets divided into *small* (batch size equal to 1) and *large* (batch size equal to 2) (Data set in format II).

comparison with that of the fitted (via MLE) exponential distribution. Note how the larger empirical quantiles are far from the fitted ones. Something similar occurs with Data set in format II, where the average, median, variation coefficient, minimum and maximum value of the inter-arrivals times are given by 0.0031, 0.0020, 1.7954, 2×10^{-5} and 0.3419 seconds. In addition, the empirical first-lag correlation coefficients of the inter-arrival times are 0.1908 and 0.2, respectively. This implies that a model capturing dependence between the arrivals may turn out to be suitable. Since arrivals occur in batches, a $BMMPP_2(2)$ and a $BMMPP_2(4)$ will be fitted to the data sets using the novel sequential fitting approach; the results shall be shown in the next section.

Results

The sequential algorithm described in Section 4 is applied to fit the teletraffic data sets. Table 3.6 shows the empirical values of a set of descriptors concerning the inter-arrivals times distribution, the batch sizes distribution and joint moments, as well as the estimated values under a $BMMPP_2(4)$ and $BMMPP_2(2)$ models for Data set in format I and II, respectively. From the first to the 10-th row, the fitted values to the characterizing moments,

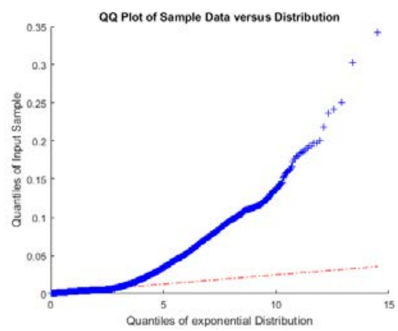


Figure 3.4: Empirical quantiles of the inter-arrival times of Data set in format I versus those of a fitted exponential distribution.

according to Theorem 2, are provided. Then, the estimated coefficients of variation, skewness and kurtosis are shown. The 14-th and 15-th rows concern the first and second moments of the batch size. Then, some descriptors related to the correlation between the inter-arrival times and the batches and the autocorrelation coefficients of the inter-event times are also depicted. The probability mass distribution of the batch size is also shown. Most of the quantities are well estimated by the models considered, with the exception of the values of $\rho_T(2)$ and $\rho_T(3)$, which are slightly underestimated.

Table 3.6 also shows a comparison between the general model proposed in the paper and a *MMPP* with i.i.d batches. For the estimation of the *MMPP*s with i.i.d batches, the sequential algorithm described in Section 4 was used, adding $q_k(\hat{y} - \hat{x}) = w_k(\hat{r} - \hat{u})$ as a restriction to each (P_k) optimization problem. In Table 3.6 it can be appreciated that in the case of Data set in format I, for which $\rho(T, B)$ and $\rho(B)$ are almost null, both estimations are quite similar. But by slightly increasing these amounts for Data set in format II the estimation using the general *BMMPP* improves over the case with independent batches. In conclusion, although the correlations or autocorrelations shown by the data are negligible, it is more reliable to adjust the general model since the computation time does not increase substantially but it does improve the quality of the adjustment in general.

The good performance of the fitted models is also supported by Figure 3.5 which depicts the fit to the empirical distribution functions of the inter-arrival times.

	Data set in format I			Data set in format II		
	Emp	Est BMMPP	Est MMPP with i.i.d. batches	Emp	Est BMMPP	Est MMPP with i.i.d. batches
μ_1	3.5625×10^{-3}	3.5625×10^{-3}	3.5625×10^{-3}	3.1428×10^{-3}	3.1428×10^{-3}	3.1428×10^{-3}
μ_2	4.7465×10^{-5}	4.7465×10^{-5}	4.7465×10^{-5}	4.1718×10^{-5}	2.0859×10^{-5}	4.1718×10^{-5}
μ_3	2.2802×10^{-6}	2.2802×10^{-6}	2.2802×10^{-6}	2.0104×10^{-6}	2.0104×10^{-6}	2.0104×10^{-6}
$\rho_T(1)$	0.1908	0.1908	0.1908	0.2	0.2	0.2
$\beta_1^{(1)}$	1.1241	1.1096	1.1040	1.8121	1.8121	1.7788
$\eta^{(1)}$	3.8663×10^{-3}	3.9189×10^{-3}	3.9329×10^{-3}	5.4932×10^{-3}	5.4932×10^{-3}	5.5905×10^{-3}
$\beta_1^{(2)}$	1.8849	1.8970	1.9019	-	-	-
$\eta^{(2)}$	6.8387×10^{-3}	6.7898×10^{-3}	6.7756×10^{-3}	-	-	-
$\beta_1^{(3)}$	1.9915	1.9935	1.9942	-	-	-
$\eta^{(3)}$	7.1082×10^{-3}	7.1037×10^{-3}	7.1044×10^{-3}	-	-	-
CV	1.6553	1.6553	1.6553	1.7954	1.7954	1.7954
$Skewness$	11.1199	11.1199	11.1199	11.1896	11.1896	11.1896
$Kurtosis$	170.3031	168.1339	168.1340	179.2370	166.8824	166.8824
β_1	1.1335	1.1162	1.1100	1.8121	1.8121	1.7788
β_2	1.4205	1.3618	1.3424	3.4364	3.4363	3.3365
$Corr(T,B)$	−0.0707	−0.0180	0	−0.0916	−0.0916	0
$\rho_B(1)$	0.0500	6.1083×10^{-3}	0	0.1037	0.1141	0
$\rho_T(2)$	0.1791	0.1146	0.1146	0.1893	0.1160	0.1160
$\rho_T(3)$	0.1278	0.0689	0.0689	0.1390	0.0673	0.0673
$P(B=1)$	0.8759	0.8904	0.8960	0.1879	0.1879	0.2212
$P(B=2)$	0.1151	0.1031	0.0891	0.8121	0.8121	0.7788
$P(B=3)$	0.0085	0.0065	0.0058	-	-	-
$P(B=4)$	4.7042×10^{-4}	1.7143×10^{-5}	1.0262×10^{-4}	-	-	-

Table 3.6: Empirical and estimated descriptors via the $BMMPP_2(4)$ (Data set in format I) and the $BMMPP_2(2)$ (Data set in format II).

Next, we focus on some quantities of interest associated to the the counting process, see Section 2.2. The top panels of Figure 3.6 show the estimated and empirical expected number of arrivals in the interval $(0, 100)$ for both Data sets in formats I and II. The bottom panels depict the estimated intervals centered on $E[N(t)] \pm kSd(N(t))$, for $k = 1, 2$, where $Sd(N(t))$ denotes

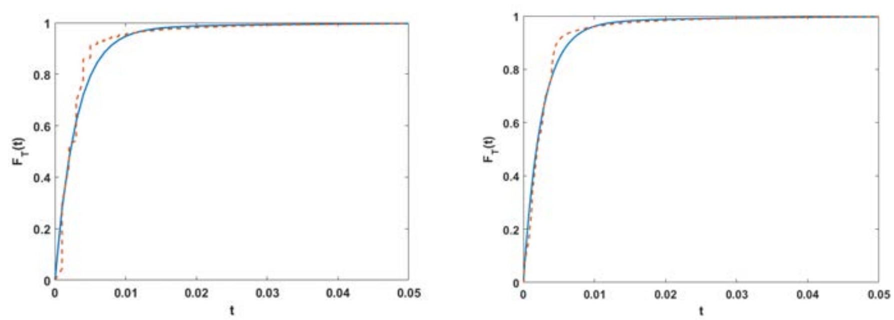


Figure 3.5: Left panel: Estimated cdf (dashed line) under the $BMMPP_2(2)$ versus the empirical cdf (solid line) of the inter-arrival times. Right panel: Estimated cdf (dashed line) under the $BMMPP_2(4)$ versus the empirical cdf (solid line) of the inter-arrival time.

the standard deviation of the number of counts, computed from (3.7). On the other hand, Figure 3.7 illustrates the estimated probabilities $p(n,t)$ as in (3.6). The left panel shows the estimated probabilities for Data set in format I for $n \in [0, 100]$ and $t = [0.1, 0.2]$. As can be observed, the sequence of functions for different values of t are bimodal, with a maximum around a high number of n, and another local maximum for a small value of n. In addition, the probability functions are not symmetric with a left tail that is longer than the right tail. Concerning Data set in format II, the left panel of Figure 3.7 shows the probabilities of the counts, for different time values and for *large* sizes. It can be seen how the variability of the variable increases with the value of t.

Finally, the queue length distribution of the $BMMPP_2(4)/M/1$ queueing system was estimated, under the assumption that the inter-arrival times of Data set in format I constitute the observed arrival process. For that, a traffic intensity $\rho = 0.5$ was set. Figure 3.8 shows the resulting tail distribution.

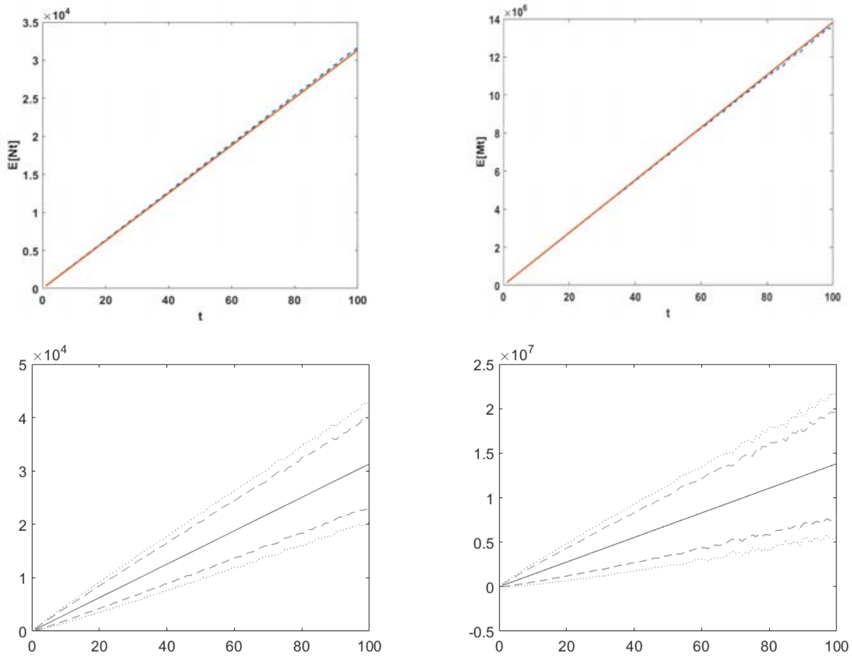


Figure 3.6: Top panels: Estimated (dashed line) and empirical (solid line) expected number of arrivals, for Data sets in formats I and II, respectively. Bottom panels: Estimated $E[N(t)]$ (solid line) and $E[N(t)] \pm kSd(N(t))$, for $k = 1$ (dashed line) and $k = 2$ (dotted line), for Data sets in formats I and II, respectively.

5 Conclusions

This paper considers the batch counterpart of the two-state Markov modulated Poisson process. The point process, noted as $BMMPP_2(K)$ turns out to be of interest in real-life contexts as reliability or queueing, since it allows for the modeling of dependent inter-event times and dependent batch sizes. The contribution of this paper is two-fold. On one hand, it is proven that the $BMMPP_2(K)$, represented by $2(K+1)$ parameters, is completely characterized in terms of a set of $2(K+1)$ moments related to the inter-event time distribution as well as to the batch size distribution. On the other

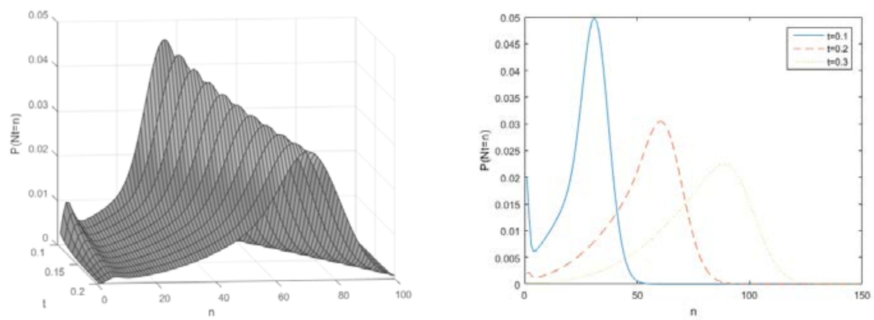


Figure 3.7: Left panel: estimated density function of the number of arrivals in different time instants for Data set in format I. Right panel: distribution of the number of *large* packets for Data set in format II for three different time instants.

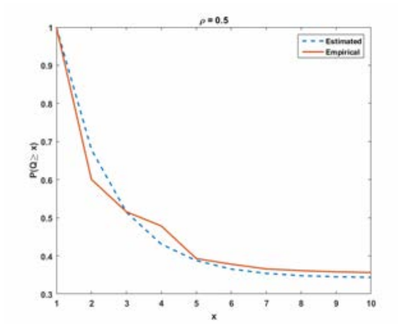


Figure 3.8: Estimated tail distribution of the queue length at departures in a $BMMPP_2(4)/M/1$ with $\rho = 0.5$ assuming that the inter-arrival times of Data set in format I constitute the observed arrival process.

hand, an inference approach for fitting real data sets based on a moments matching method is described. The method involves solving, in an iterative way, $K-1$ optimization problems with two unknowns, yielding an efficient and tractable algorithm. The performance of the novel inference technique is illustrated using both simulated and a real teletraffic trace, for which the queue length distribution at departures in a $BMMPP_2(K)/M/1$ queue are estimated. The method is also compared to the classic EM algorithm

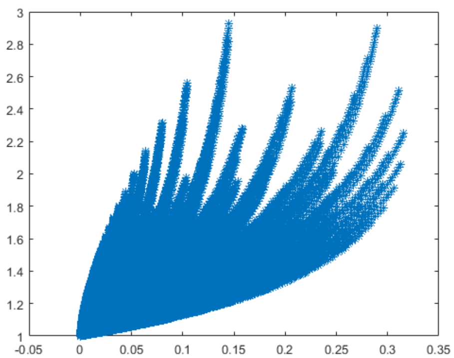


Figure 3.9: Scatter plot of $\rho_T(1)$ versus the coefficient of variation of the inter-event times from 700000 simulated $BMMPP$s

which has been considered by previous studies dealing with inference for the $BMAP$. The results show that the novel approach turns out to be faster and less dependent on starting points than the EM algorithm.

Prospects regarding this work concern both applied and theoretical issues. In the first case, given that higher order $BMMPP_m(K)$ are expected to show more versatility for modeling purposes (Rodríguez et al., 2016b), it is of interest to develop inference methods in these cases. A moment-matching approach similar to the one proposed in this paper could be considered for higher order $BMMPP$s (which are known to be identifiable, Yera et al. (2019)). However, the set of moments characterizing $BMMPP_m(K)$ processes is still unknown when $m \geq 3$. From a theoretical viewpoint, a challenging problem to be considered is related to the correlation structures (of both inter-event times and batch sizes) of the $BMMPP_m(K)$, for $m \geq 3$. Similar approaches such as in Ramírez-Cobo and Carrizosa (2012); Rodríguez et al. (2016b) shall be taken into account to address this issue. Finally, another theoretical problem that needs to be examined in more detail refers to the sample sizes required for the estimation method. In this direction, Ramírez-Cobo et al. (2017) suggest that the values of the coefficient of variation of the inter-event times (CV) and the first-lag autocorrelation coefficient ($\rho_T(1)$) are positively correlated. This would imply that the required sample size should increase with the value of the correlation between consecutive events. From Figure 3.9, it can be seen that even though there exist processes for which the CV is high and the value of $\rho_T(1)$ is low, it is true that high values of $\rho_T(1)$ seem to be linked to values of the CV larger than a lower bound ($CV \sim 1.6$). On the contrary, if $\rho_T(1)$ is very close to zero, then the CV seems to be closer to 1. This problem is an open question that, together with the previous issues, will be undertaken in future work.

Acknowledgments

The authors would like to thank the anonymous referees whose comments and suggestions improved the understanding of this paper. Research partially supported by research grants and projects MTM2015-65915-R and ECO2015-66593-P (Ministerio de Economía y Competitividad, Spain) and P11-FQM-7603, FQM-329 (Junta de Andalucía, Spain) and Fundación BBVA.

Appendix A: Proof of the 3.5

Using the Laplace-Stieltjes transform (LST) of the n first inter-event times and batch sizes of a stationary $BMAP_2(K)$ given in Rodríguez et al. (2016c), then $E[TB]$ is found as

$$\begin{aligned}
\eta = E[TB] &= E\left[\frac{\partial(-e^{-sT}z^B)}{\partial s \partial z}\right]\Bigg|_{s=0,z=1} = -\frac{\partial f^*_{T,B}(s,z)}{\partial s \partial z}\Bigg|_{s=0,z=1} \\
&= -\frac{\partial}{\partial z}\left[\frac{\partial}{\partial s}\left[\phi(sI - \boldsymbol{D_0})^{-1}\left(\sum_{k=1}^{K} z^k \boldsymbol{D_k}\right)\boldsymbol{e}\right]\right]\Bigg|_{s=0,z=1} \\
&= \left[\phi(sI - \boldsymbol{D_0})^{-2}\left(\sum_{k=1}^{K} k z^{k-1} \boldsymbol{D_k}\right)\boldsymbol{e}\right]\Bigg|_{s=0,z=1} \\
&= \phi(-\boldsymbol{D_0})^{-2}\boldsymbol{D_1^*}\boldsymbol{e}
\end{aligned}$$

Appendix B: $cor(T;B)$ and ρ_B are zero for a *MMPP* with i.i.d. batches (Remark 1)

Proposition 6. *Let $\mathcal{B} = \{\boldsymbol{D_0}, p_1\boldsymbol{\Delta(\delta)}, ..., p_K\boldsymbol{\Delta(\delta)}\}$ be a stationary* MMPP *with i.i.d. batches. Then,*

(i) the autocorrelation function of the batch sizes, $\rho_B(l) = 0$, for all $l \geq 1$

(ii) the covariance between T and B is equal to zero.

Proof. Consider the first-lag autocorrelation coefficient given by

$$\rho_B(1) = \frac{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*e} - (\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*e})^2}{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_2^*e} - (\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}D_1^*\boldsymbol{e})^2}. \tag{3.20}$$

Using that $\boldsymbol{D_k} = p_k\boldsymbol{\Delta(\delta)}$ for $k = 1, ..., K$, (3.20) can be rewritten as

$$\begin{aligned}
\rho_B(1) &= \frac{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)e}}{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K k^2p_k\right)\boldsymbol{\Delta(\delta)e} - \left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)e}\right]^2} \\
&\quad - \frac{\left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)e}\right]^2}{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K k^2p_k\right)\boldsymbol{\Delta(\delta)e} - \left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)e}\right]^2} \\
&= \frac{\left(\sum_{k=1}^K kp_k\right)^2\left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} - (\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e})^2\right]}{\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K k^2p_k\right)\boldsymbol{\Delta(\delta)e} - \left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\left(\sum_{k=1}^K kp_k\right)\boldsymbol{\Delta(\delta)e}\right]^2} \\
&= \frac{\left(\sum_{k=1}^K kp_k\right)^2\left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} - 1\right]}{\left(\sum_{k=1}^K k^2p_k\right)\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} - \left(\sum_{k=1}^K kp_k\right)^2\left[\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e}\right]^2} \qquad (3.21) \\
&= \frac{\left(\sum_{k=1}^K kp_k\right)^2[1-1]}{\left(\sum_{k=1}^K k^2p_k\right) - \left(\sum_{k=1}^K kp_k\right)^2} \\
&= 0.
\end{aligned}$$

In (3.21) it is used that $(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} = \boldsymbol{e}$ and $\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} = 1$, which can be derived from $\boldsymbol{Qe}=\boldsymbol{0}$ as follow

$$\begin{aligned}
(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e} &= (-\boldsymbol{D_0})^{-1}(\boldsymbol{Q}-\boldsymbol{D_0})\boldsymbol{e} \\
&= (-\boldsymbol{D_0})^{-1}\boldsymbol{Qe}-(-\boldsymbol{D_0})^{-1}\boldsymbol{D_0e} \\
&= (-\boldsymbol{D_0})^{-1}\boldsymbol{Qe}+\boldsymbol{e} \\
&= \boldsymbol{e}, \qquad (3.22)
\end{aligned}$$

and consequently,

$$\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e}=\boldsymbol{\phi e}=1. \qquad (3.23)$$

In Rodríguez et al. (2016b), it is proven that for a general $BMAP_2(K)$, the auto-correlation function decreases in absolute value ($|\rho_B(l)|\geq|\rho_B(l{+}1)|$ for all $l\geq 1$). Therefore, if $\rho_B(1)=0$ implies $\rho_B(l)=0$ for all $l\geq 1$.

For the second proposition, using the expresion for η, μ_1, β_1 and (3.22)-(3.23), the nullity of $Cov(T,B)$ is proven as follow

$$\begin{aligned}
Cov(T,B) &= \phi(-\boldsymbol{D_0})^{-2}\boldsymbol{D_1^*e}-\mu_1\beta_1 \\
&= \phi(-\boldsymbol{D_0})^{-2}\boldsymbol{D_1^*e}-\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1^*e} \\
&= \phi(-\boldsymbol{D_0})^{-2}\left(\sum_{k=1}^{K}kp_k\right)\boldsymbol{\Delta(\delta)e}-\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e\phi}(-\boldsymbol{D_0})^{-1} \\
&\quad\times\left(\sum_{k=1}^{K}kp_k\right)\boldsymbol{\Delta(\delta)e} \\
&= \left(\sum_{k=1}^{K}kp_k\right)\left[\phi(-\boldsymbol{D_0})^{-2}\boldsymbol{\Delta(\delta)e}-\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta(\delta)e}\right] \\
&= \left(\sum_{k=1}^{K}kp_k\right)\left[\phi(-\boldsymbol{D_0})^{-1}\boldsymbol{e}-\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e}\right] \\
&= 0.
\end{aligned}$$

□

Appendix C: G_0 and G_1 in Lemma 6 are well defined

First, denote as G_{ij0} the $(i,j)-$th element of

$$G_0 = \frac{1}{a\zeta_1 - b\zeta_2}\begin{pmatrix} \zeta_1\zeta_2 - a\zeta_1^2 - a\zeta_1\zeta_2 + ab\zeta_1\zeta_2 & -a^2\zeta_1^2 + a\zeta_1^2 + \zeta_2 a\zeta_1 - \zeta_2\zeta_1 \\ (\zeta_1 - b\zeta_2)(\zeta_2 - b\zeta_2) & -\zeta_1\zeta_2 + b\zeta_2^2 + b\zeta_1\zeta_2 - ab\zeta_1\lambda_2 \end{pmatrix}$$

and, in analogous way, let G_{ij1} the $(i,j)-$th element of

$$G_1 = Diag(G_1^c) = \begin{pmatrix} a\zeta_1 & 0 \\ 0 & b\zeta_2 \end{pmatrix}.$$

The aim is to prove that G_0 and G_1 are indeed matrices defining a *MMPP*, that is.

$$\begin{cases} G_{ii0} < 0 \quad i = \{1,2\} \\ 0 \leq G_{ij0} < \infty \quad i \neq j, i, \quad j = \{1,2\} \\ (G_0 + G_1)\boldsymbol{e} = \mathbf{0} \end{cases}$$

From the definition of the $MMPP_2$, the cannonical form for the *MAP* and the initial assumption $x + y \geq r + u$, it is not difficult to see that $a\zeta_1$, $b\zeta_2$ and $b\zeta_2 - a\zeta_1$ are non-negative, therefore

$$\begin{aligned} G_{120} + G_{111} &= \frac{-a^2\zeta_1^2 + a\zeta_1^2 + \zeta_2 a\zeta_1 - \zeta_2\zeta_1}{a\zeta_1 - b\zeta_2} + a\zeta_1 \\ &= \frac{-a^2\zeta_1^2 + a\zeta_1^2 + \zeta_2 a\zeta_1 - \zeta_2\zeta_1 + a\zeta_1(a\zeta_1 - b\zeta_2)}{a\zeta_1 - b\zeta_2} \\ &= \frac{a\zeta_1^2 + \zeta_2 a\zeta_1 - \zeta_2\zeta_1 - a\zeta_1 b\zeta_2}{a\zeta_1 - b\zeta_2} \\ &= \frac{\zeta_1[a\zeta_1 - \zeta_2(1 - a + ab)]}{a\zeta_1 - b\zeta_2} \end{aligned} \tag{3.24}$$

$$\geq \quad 0$$

Note that $1 - a + ab > b$, hence both the numerator and the denominator in (3.24) are non-positive.

On the other hand,

$$G_{110} + G_{120} = \frac{ab\zeta_1\zeta_2 - a^2\zeta_1^2}{a\zeta_1 - b\zeta_2} = \frac{a\zeta_1(b\zeta_2 - a\zeta_1)}{a\zeta_1 - b\zeta_2} = -a\zeta_1 = -G_{111}$$

Therefore $G_{110} + G_{120} + G_{111} = 0$, $G_{110} \leq 0$ and $G_{120} \geq 0$.

Similarly, for the second line we have

$$\begin{aligned} G_{210} + G_{221} &= \frac{(\zeta_1 - b\zeta_2)(\zeta_2 - b\zeta_2)}{a\zeta_1 - b\zeta_2} + b\zeta_2 \\ &= \frac{(\zeta_1 - b\zeta_2)(\zeta_2 - b\zeta_2) + b\zeta_2(a\zeta_1 - b\zeta_2)}{a\zeta_1 - b\zeta_2} \\ &= \frac{\zeta_2\zeta_1 + a\zeta_1 b\zeta_2 - b\zeta_2^2 - \zeta_2 b\zeta_1}{a\zeta_1 - b\zeta_2} \\ &= \frac{\zeta_2[\zeta_1(1 - b + ab) - b\zeta_2]}{a\zeta_1 - b\zeta_2} \qquad (3.25) \\ &\geq 0 \end{aligned}$$

Note that $a < 1 - b + ab$, hence both the numerator and the denominator in (3.25) are non-positive, which completes the proof.

References

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

Andersen, A. T. and Nielsen, B. F. (2002). On the use of second-order descriptors to predict queueing behavior of maps. *Naval Research Logistics (NRL)*, 49(4):391–409.

Arts, J. (2017). A multi-item approach to repairable stocking and expediting in a fluctuating demand environment. *European Journal of Operational Research*, 256(1):102–115.

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

Banerjee, A., Gupta, U., 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.

Banik, A. and Chaudhry, M. (2016). Efficient computational analysis of stationary probabilities for the queueing system $BMAP/G/1/N$ with or without vacation (s). *INFORMS Journal on Computing*, 29(1):140–151.

Bean, N. and Green, D. (1999). When is a MAP poisson? *Mathematical and Computer Modelling*, 82:127–142.

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.

Breuer, L. (2002). An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, 112:123–138.

Buchholz, P. and Kriege, J. (2017). Fitting correlated arrival and service times and related queueing performance. *Queueing Systems*, 85(3-4):337–359.

Carrizosa, E. and Ramírez-Cobo, P. (2014). Maximum likelihood estimation in the two-state Markovian arrival process. *arXiv preprint arXiv:1401.3105*, Working paper.

Casale, G., Sansottera, A., and Cremonesi, P. (2016). Compact markov modulated models for multiclass trace fitting. *European Journal of Operational Research*, 255(3):822–833.

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

Chakravarthy, S. R. (2010). Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science*.

Eum, S., Harris, R., and Atov, I. (2007). A matching model for MAP-2 using moments of the counting process. In *Proceedings of the International Network Optimization Conference, INOC 2007*, Spa, Belgium.

Fearnhead, P. and Sherlock, C. (2006). An exact Gibbs sampler for the Markov modulated poisson process. *Journal of the Royal Statististical Society: Series B*, 65(5):767–784.

Ghosh, S. and Banik, A. (2017). An algorithmic analysis of the $BMAP/MSP/1$ generalized processor-sharing queue. *Computers & Operations Research*, 79:1–11.

Green, D. (1998). *$MAP/PH/1$ departure processes*. PhD thesis, School of Applied Mathematics, University of Adelaide, South Australia.

He, Q.-M. and Zhang, H. (2006). PH-invariant polytopes and coxian representations of phase type distributions. *Stochastic Models*, 22(3):383–409.

He, Q.-M. and Zhang, H. (2008). An algorithm for computing minimal coxian representations. *INFORMS Journal on Computing*, 20:179–190.

He, Q.-M. and Zhang, H. (2009). Coxian representations of generalized Erlang distributions. *Acta Mathematicae Applicatae Sinica. English Series*, 25:489–502.

Heffes, H. and Lucantoni, D. (1986). A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, 4:856–868.

Horváth, M., Buchholz, P., and Telek, M. (2005). A map fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. *IEEE/ Second International Conference on the Quantitative Evaluation of Systems (QEST'05)*, pages 124–133.

Kang, S. H. and Sung, D. K. (1995). Two-state $MMPP$ modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times. In *Proceedings of GLOBECOM'95*, volume 2, pages 1422–1426. IEEE.

Klemm, A., Lindemann, C., and Lohmann, M. (2003). Modeling IP traffic using Batch Markovian Arrival Process. *Performance Evaluation*, 54(2):149–173.

Kriege, J. and Buchholz, P. (2011). Correlated phase-type distributed random numbers as input models for simulations. *Performance Evaluation*, 68(11):1247–1260.

Landon, J., Özekici, S., and Soyer, R. (2013). A markov modulated poisson model for software reliability. *European Journal of Operational Research*, 229(2):404–410.

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM.

Li, M., Chen, W., and Han, L. (2010). Correlation matching method for the weak stationarity test of lrd traffic. *Telecommunication Systems*, 43(3-4):181–195.

Liu, B., Cui, L., Wen, Y., and Shen, J. (2015). A cold standby repairable system with working vacations and vacation interruption following markovian arrival process. *Reliability Engineering & System Safety*, 142:1–8.

Lucantoni, D. (1991). New results for the single server queue with a Batch Markovian Arrival Process. *Stochastic Models*, 7:1–46.

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

Lucantoni, D., Choudhury, G., and Whitt, W. (1994). The transient $BMAP/G/1$ queue. *Stochastic Models*, 10:145–182.

Lucantoni, D., Meier-Hellstern, K., and Neuts, M. (1990). A single-server queue with server vacations and a class of nonrenewal arrival processes. *Advances in Applied Probability*, 22:676–705.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2015). A reliability system under cumulative shocks governed by a bmap. *Applied Mathematical Modelling*, 39(23-24):7620–7629.

Narayana, S. and Neuts, M. (1992). The first two moment matrices of the counts for the Markovian arrival process. *Communications in statistics. Stochastic models*, 8(3):459–477.

Nasr, W. W., Charanek, A., and Maddah, B. (2018). Map fitting by count and inter-arrival moment matching. *Stochastic Models*, pages 1–29.

Neuts, M. and Li, J. (1997). *An algorithm for the $P(n,t)$ matrices of a continuous BMAP*, volume 183 of *Lectures notes in Pure and Applied Mathematics*, pages 7–19. Srinivas R. Chakravarthy and Attahiru, S. Alfa, editors. NY: Marcel Dekker, Inc.

Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779.

Okamura, H., Dohi, T., and Trivedi, K. (2011). A refined em algorithm for ph distributions. *Performance Evaluation*, 68(10):938–954.

Ramaswami, V. (1990). From the matrix-geometric to the matrix-exponential. *Queueing Systems*, 6:229–260.

Ramírez-Cobo, P. and Carrizosa, E. (2012). A note on the dependence structure of the two-state Markovian arrival process. *Journal of Applied Probability*, 49:295–302.

Ramírez-Cobo, P. and Lillo, R. (2012). New results about weakly equivalent MAP_2 and MAP_3 processes. *Methodology and Computing in Applied Probability*, 14(3):421–444.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2008). Bayesian analysis of a queueing system with a long-tailed arrival process. *Communications in StatisticsSimulation and Computation*, 37(4):697–712.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010). Nonidentifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47(3):630–649.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2017). Bayesian analysis of the stationary MAP_2. *Bayesian Analysis*, 12(4):1163–1194.

Ramírez-Cobo, P., Lillo, R. E., Wilson, S., and Wiper, M. P. (2010). Bayesian inference for Double Pareto lognormal queues. *Annals of Applied Statistics*, 4(3):1533–1557.

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

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 and System Safety*, 134:126–133.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016a). Analytical issues regarding the lack of identifiability of the non-stationary MAP_2. *Performance Evaluation*, 102:1–20.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016b). Dependence patterns for modeling simultaneous events. *Reliability Engineering and System Safety*, 154:19–30.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016c). Nonidentifiability of the two-state $BMAP$. *Methodology and Computing in Applied Probability*, 18(1):81–106.

Rydén, T. (1994). Parameter estimation for Markov modulated Poisson processes. *Stochastic Models*, 10(4):795–829.

Rydén, T. (1996a). An EM algorithm for estimation in Markov modulated Poisson processes. *Computational Statistics & Data Analysis*, 21(4):431–447.

Rydén, T. (1996b). 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.

Scott, S. (1999). Bayesian analysis of the two state markov modulated poisson process. *Journal of Computational and Graphical Statistics*, 8(3):662–670.

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.

Sikdar, K. and Samanta, S. (2016). Analysis of a finite buffer variable batch service queue with batch markovian arrival process and servers vacation. *Opsearch*, 53(3):553–583.

Singh, G., Gupta, U., and Chaudhry, M. (2016). Detailed computational analysis of queueing-time distributions of the $BMAP/G/1$ queue using roots. *Journal of Applied Probability*, 53(4):1078–1097.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019). Findings about the two-state BMMPP for modeling point processes in reliability and queueing systems. *Applied Stochastic Models in Business and Industry*, 35(2):177–190.

Part III

Bivariate Markov modulated Poisson processes

CHAPTER 4

A bivariate two-state Markov modulated Poisson process for failure modeling

Yoel G. Yera[1], Rosa E. Lillo[1,2], Pepa Ramírez-Cobo[3], Fabrizio Ruggeri[4] and Bo F. Nielsen[5]

[1] Department of Statistics, Universidad Carlos III de Madrid, Spain
[2] UC3M-BS Institute of Financial Big Data Getafe, Madrid, Spain
[3] Department of Statistics and Operations Research, Universidad de Cádiz, Spain
[4] IMATI-CNR, Institute of Applied Mathematics and Information Technologies, Milano, Italy
[5] Department of Applied Mathematics and Computer Science. Technical University of Denmark, Kongens Lyngby, Denmark

Abstract

This paper presents an extension to the two-dimensional case of the Markov modulated Poisson process (*MMPP*), motivated by real failure data in a two-dimensional context. The one-dimensional *MMPP* has been proposed for the modeling of dependent and non-exponential inter-event times (in contexts as queuing, risk or reliability, among others). The novel two-dimensional *MMPP* allows for dependence among the two sequences of inter-event times, while at the same time preserves the *MMPP* properties, marginally. Such generalization is based on the Marshall-Olkin exponential distribution. Inference is undertaken for the new process through a method combining a matching moments approach and an ABC algorithm. The performance of the method is shown on simulated and real datasets representing failures of a public transport company.

Keywords: Stochastic processes. Markov modulated Poisson process (*MMPP*). Moments matching method. Teletraffic data.

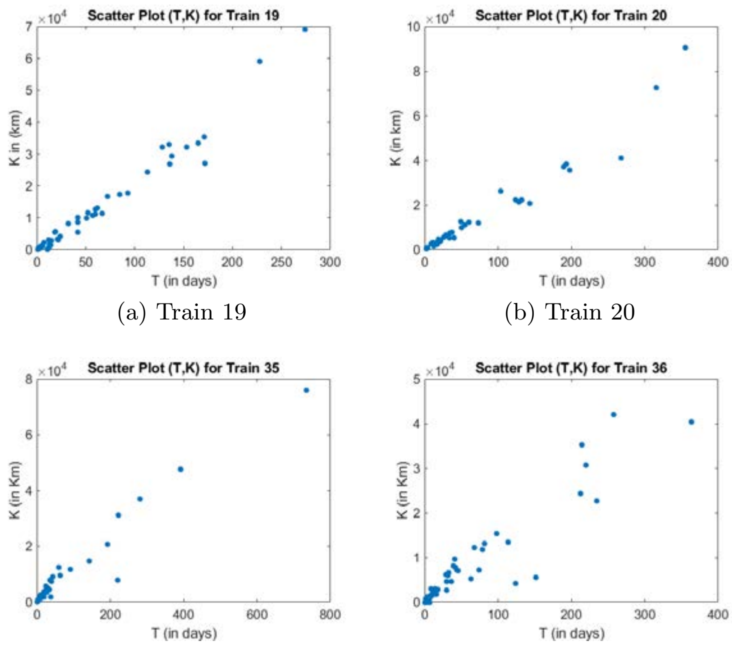
1 Introduction

Transportation means are essential in every day life and it is crucial to know if there is any failure, what its nature is and how to prevent it. Reliability studies associated with airplanes, automobiles or ships have appeared in several academic articles, see for example Goel and Gupta (1984), Sheikh et al. (1996), Al-Garni et al. (1999), Zhang and Liu (2002), Zhang et al. (2005), Karim (2008) Fang and Das (2005), Ivanov (2009) and Decò et al. (2012). With the rapid development of the railway industry and especially with high-speed lines, the safety and reliability of the train system has attracted increasing attention. Over the years, more and more safety and operational reliability analyses of train traffic have been carried out. For example, Eliashberg et al. (1997) develop a model that describes failures by two scales, for example time and mileage, and applies that model to calculate the optimal reserve that a manufacturer of locomotive traction motors should have to cover the warranty of possible breakages; Bearfield and Marsh (2005) use event tree analysis and study train derailments; Pievatolo et al. (2003), Ruggeri (2006) and Pievatolo and Ruggeri (2010) use non-homogeneous Poisson processes to assess if the reliability of underground train doors was compliant with the contract signed by a manufacturer and a transportation company; Jafarian and Rezvani (2012) present a study on the evaluation of the railway safety risks using the fuzzy fault tree analysis and Navas et al. (2017) use various models to estimate the reliability in railway repairable systems. In this article we consider the same dataset as in

Pievatolo et al. (2003), which represents the records of the failures of trains for about 8 years. In particular, the data records failures from 40 underground trains, which were delivered to an European transportation company between November 1989 and March 1991 and all of them were put in service from 20th March 1990 to 20th July 1992. Failure monitoring ended on 31st December 1998. When a failure took place, both the reading of the odometer (which quantifies the number of kilometres covered) and the date of the failure were recorded, which suggests the need of a two-dimensional stochastic process to model the data. This data set was motivated by the need for detecting as soon as possible any unexpected poor reliability of the component (as agreed in the contract signed with the manufacturer). In Pievatolo et al. (2003) a Poisson process with double measurements (time and kilometres) related by a gamma process is proposed. Later, the same problem was addressed in Pievatolo and Ruggeri (2010), where the authors use again a Poisson process, but now considering a bivariate intensity function.

An exploratory analysis of the dataset is considered next. On one hand, the scatter plots of four trains shown in Figure 4.1 evidence a high linear dependence between the inter-failure times and distances (intra-dependence). On the other hand, Figure 4.2 shows the first-lag autocorrelation coefficient of the inter-failure times versus those of the distances between failures. From Figure 4.2, a non-negligible inter-dependence in the sequence is observed. Another important feature of this data is that, neither the inter-failure times, nor the inter-failure distances, seem to come from an exponential distribution, as it can be seen from Figure 1.5. Therefore, it would be desirable to find a bivariate process, versatile enough to allow for non-exponential spacing (both in time and distance) between failures, as well as able to model both intra and inter-dependence in the observed sequences. Therefore, the aim of this paper is to present a tractable, analytical model able to jointly fit the data set and from which performance quantities of interest are derived in straightforward way. In analogous circumstances but in univariate case, some authors recommend the use of the Markovian arrival process (*MAP*), see for example Ramírez-Cobo et al. (2017), Buchholz (2003), Klemm et al. (2003), Ramírez-Cobo et al. (2008), Ramírez-Cobo and Carrizosa (2012), Rodríguez et al. (2016a) and Rodríguez et al. (2015). However, it is known that *MAP*s cannot be identified in a unique way, which is inconvenient for their statistical estimation, see for example Bodrog et al. (2008), Ramírez-Cobo et al. (2010) and Rodríguez et al. (2016b). Hence, the identifiable subclass of Markov-modulated Poisson process (*MMPP*s) is preferable. For more details on the identifiability of the *MMPP* and its effect on the estimation, Rydén (1996a), Yera et al. (2019a) and Yera et al. (2019b) can be reviewed. In this paper we propose a stochastic process that maintains the same good analytic properties of the *MMPP*, namely, a matrix representation, non-exponential distributions of the inter-failures times (and distances), dependence among failures and able to model correlations between the two observed sequences. Several models have been already proposed in the literature for modeling two-dimensional inter-event sequences. For example, Griffiths and Milne (1978) and Griffiths et al. (1979) propose the bivariate Poisson process and Assaf et al. (1984) and Badescu et al. (2009) consider the bivariate phase-type distribution. Pievatolo et al. (2003) and Pievatolo and Ruggeri (2010) make use of a non-homogeneous Poisson processes with a double scale to study the database analyzed in this paper. Nadar and Kızılaslan (2015) and Yuan (2018) use different types of bivariate Weibull distributions, while Kızılaslan and Nadar (2018) make use of a bivariate Kumaraswamy distribution. On the other hand, Cai and Li (2005) and Zadeh and Bilodeau (2013) use bivariate phase type distributions to solve problems of risk and insurance loss models respectively. However, all these approaches to the problem of modeling events in two dimensions use processes in which the times between events are independent and as can be seen in Figure 2 this is a starting hypothesis that does not have the data on which this work is based. Therefore, our proposal is the extension of the *MMPP* to its bivariate counterpart, which, to the best of our knowledge, has not been done previously in the literature. For simplicity, in this paper we will focus on the $MMPP_2$.

The goal of the paper is twofold. The first is to propose an extension of the $MMPP_2$ to the bivariate case, in such a way that the statistical features previously commented are well modeled. Some theoretical properties of this



(a) Train 19

(b) Train 20

(c) Train 35

(d) Train 36

Figure 4.1: Linear relationship between inter-failure times and distances for four trains

novel process such as the identifiability will also be studied. The identifiability of the process is a crucial property to be taken into consideration in order to develop an estimation method, since it determines the possible multimodality of the likelihood function. Therefore, from a theoretical view point, the identifiability of the novel process will be addressed in this paper. The second main objective is to propose an estimation method for the novel process. As it will be seen, it will be divided into two steps. The first one is based on a matching moments approach and it provides an estimate for some parameters, those linked to the marginal processes. The second step, which consists in an ABC algorithm in terms of some joint moments, generates estimates for the remaining parameters.

After a brief review of the $MMPP_2$ and the bivariate Marshall-Olkin Exponential distribution in Section 2.1, Section 2.2 describes the two-

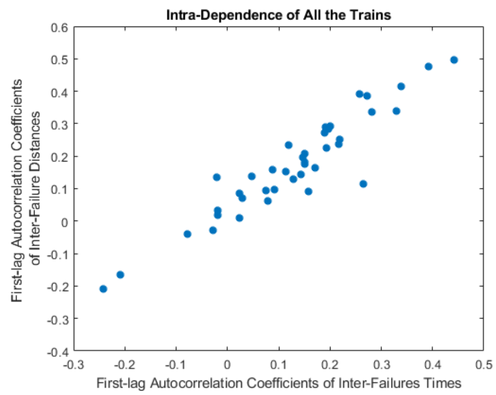


Figure 4.2: Fist-lag autocorrelation coefficient of the inter-failure times and distances for each train

dimensional version of the $MMPP_2$. Section 2.3 addresses the issue of the identifiability of the novel process, while a matrix representation for the process is provided in Section 2.4. The statistical inference part of the process is addressed in Section 3. In particular, Section 3.1 describes the fitting algorithm. Section 3.2 illustrates the performance of the algorithm on simulated traces. In Section 3.3 a real application of the novel approach is considered to model the dataset related to train failures described at the beginning of Section 1. In the numerical analyses, the estimation of some conditional probabilities of the bivariate processes are also considered. Finally, Section 4 presents conclusions and delineates possible directions for future research.

2 A bivariate two-state Markov Modulated Poisson Process

In this section the two-state bivariate Markov Modulated Poisson Process is introduced. This model can be considered as an extension of the classical two-state Markov Modulated Poisson Process ($MMPP_2$). This section is divided into four parts to facilitate the reading and understanding of the results. The first one is devoted to review the $MMPP_2$ as well as the Marshall-Olkin bivariate exponential distribution on which the novel bivariate model is based. Section 2.2 formally introduces the new bivariate process where an algorithm to simulate traces is also included. Section 2.4 provides a matrix representation of the process as well as some theoretical quantities of interest. Finally, Section 2.3 considers the problem of identifiability of the new bivariate model.

2.1 Preliminaries

The $MMPP_2$ is governed by a two-states underlying Markov process $J(t)$ with infinitesimal generator $\boldsymbol{Q}$ on $\{1, 2\}$. The $MMPP_2$ is also frequently referred to as a Switched Poisson Process (SPP), see e.g. van Hoorn and Seelen (1983). Then, at the end of an exponentially distributed sojourn time in state i, with mean $1/\lambda_i$, two possibilities can occur. First, with probability a if $i = 1$ (b if $i = 2$), no event occurs and the system enters into the other state $j \neq i$. Second, with probability $1-a$ if $i = 1$ ($1-b$ if $i = 2$), an event is produced and the system continues in the same state.

The $MMPP_2$ can be characterized in terms of rate (or intensity) matrices $\{\boldsymbol{D_0}, \boldsymbol{D_1}\}$ where

$$\boldsymbol{D_0} = \begin{pmatrix} -\lambda_1 & \lambda_1 a \\ \lambda_2 b & -\lambda_2 \end{pmatrix}, \quad \boldsymbol{D_1} = \begin{pmatrix} \lambda_1(1-a) & 0 \\ 0 & \lambda_2(1-b) \end{pmatrix}. \tag{4.1}$$

This definition of the rate matrices implies that $\boldsymbol{Q} = \boldsymbol{D_0} + \boldsymbol{D_1}$.

For a better understanding of the model's behaviour, Figure 4.3 illustrates a realization of the $MMPP_2$.

If T_n denotes the time between the $(n-1)$-th and n-th events, the inter-event times, T_ns, are phase-type distributed with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}\}$, where $\boldsymbol{\phi}$ is the stationary probability vector associated with $\boldsymbol{P^*} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}$, computed as $\boldsymbol{\phi} = (\boldsymbol{\pi}\boldsymbol{D_1}\boldsymbol{e})^{-1}\boldsymbol{\pi}\boldsymbol{D_1}$ (see Latouche and Ramaswami (1999) and Chakravarthy (2010)), where $\boldsymbol{\pi}$ is the stationary probability vector of $\boldsymbol{Q}$ and $\boldsymbol{e}$ is a vector of ones. This implies that the cumulative distribution function of T_n is given by

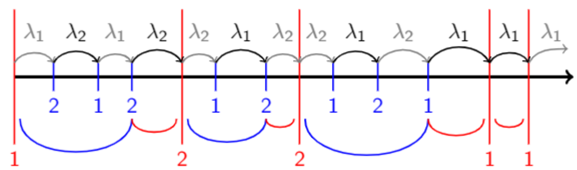


Figure 4.3: Transition diagram for the $MMPP_2$. The blue line corresponds to transitions without events, governed by $\boldsymbol{D_0}$, and the red lines correspond to transitions with events, governed by $\boldsymbol{D_1}$.

$$F_{T_n}(t) = 1 - \boldsymbol{\alpha_n} e^{\boldsymbol{D_0} t} \boldsymbol{e}, \tag{4.2}$$

where $\boldsymbol{\alpha_n} = \boldsymbol{\phi}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}]^{(n-1)} = \boldsymbol{\phi}$, since the process is stationary $\boldsymbol{\phi}$ is defined to satisfy $\boldsymbol{\phi}[(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}] = \boldsymbol{\phi}$. It is easy to check that for the case of the $MMPP_2$, $\boldsymbol{\phi}$ only depends on the probabilities associated with the underlying process:

$$\boldsymbol{\phi} = (\phi_1, \phi_2) = \left(\frac{b(1-a)}{b(1-a)+a(1-b)}, \frac{a(1-b)}{b(1-a)+a(1-b)} \right). \tag{4.3}$$

On the other hand, the moments of T_n in the stationary case are given by

$$\mu_T(r) = E(T^r) = r!\boldsymbol{\phi}(-\boldsymbol{D_0})^{-r}\boldsymbol{e}, \quad \text{for } r \geq 1, \tag{4.4}$$

and the auto-correlation function of the sequence of inter-failure times is

$$\rho_T(l) = \rho(T_1, T_{l+1}) = \frac{\pi[(-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}]^l(-\boldsymbol{D_0})^{-1}\boldsymbol{e} - \mu_T(1)}{2(\boldsymbol{\pi}(-\boldsymbol{D_0})^{-1}\boldsymbol{e} - \mu_T(1))}, \quad \text{for } l > 0. \tag{4.5}$$

Further details of the properties of the $MMPP$ can be found, for example, in Fischer and Meier-Hellstern (1993), Rydén (1996a) or Yera et al. (2019a).

The construction of the bivariate $MMPP$ proposed in this article is based on maintaining the same underlying structure of the univariate case, but replacing the exponential univariate distribution with a bivariate exponential distribution.

As it is known in the literature, there are several options to define a bivariate exponential distribution and in principle any would serve for the purpose of simulating the bivariate MMPP. In this paper we have applied the bivariate Marshall-Olkin exponential distribution, which is a bivariate distribution that fits into the MPH^* framework proposed in Kulkarni (1989), see also Chapter 8 in Bladt and Nielsen (2017). The distribution is originally defined in Marshall and Olkin (1967) as follows:

Definition 10. *Let X and Y be positive continuous random variables. Then X and Y are distributed according to the bivariate exponential distribution (BVE) with parameters $\lambda_1, \lambda_2, \lambda_3$, noted as $(X, Y) \sim BVE(\lambda_1, \lambda_2, \lambda_3)$ if*

$$\bar{F}(x, y) = P(X > x, Y > y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)\},$$

where $\lambda_1, \lambda_2 > 0$ and $\lambda_3 \geq 0$.

In Marshall and Olkin (1967) basic properties of the BVE are provided. For example, the density function is given by

$$f(x, y) = \begin{cases} \lambda_1 \gamma_k \exp\{-\lambda_1 x - \gamma_k y\} & \text{if } x < y \\ \lambda_2 \gamma_t \exp\{-\gamma_t x - \lambda_2 y\} & \text{if } x > y \\ \lambda_3 \exp\{-(\lambda_1 + \lambda_2 + \lambda_3) x\} & \text{if } x = y \end{cases}$$

The measure has a singular decomposition with a part that is absolutely continuous with respect to the two-dimensional Lebesgue measure in the first quadrant and a measure on the half line $x = y$ in the first quadrant. From the definition it can be shown that X and Y follow an exponential distribution with means $1/\gamma_t$ and $1/\gamma_k$ respectively, where $\gamma_t = \lambda_1 + \lambda_3$ and $\gamma_k = \lambda_2 + \lambda_3$. Marshall and Olkin (1967) also deduce the Laplace transform, and consequently, a formula for the joint moments

$$E(XY) = \frac{1}{\lambda_1 + \lambda_2 + \lambda_3} \left(\frac{1}{\gamma_t} + \frac{1}{\gamma_k} \right) \tag{4.6}$$

and the correlation between X and Y

$$\rho(X,Y) = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3},$$

respectively. The selection of the Marshall-Olkin BVE is also motivated by the fact that the correlation between the two sequences of events generated by this distribution, ρ, is non-negative as it is the case of the empirical correlation observed in the train failures dataset.

In Kulkarni (1989) and Bladt and Nielsen (2017) the bivariate Marshall-Olkin exponential distribution is represented as a multivariate phase-type distribution and denoted as MPH^*. This representation is given by an initial probability vector and two matrices $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ as follow:

$$\left((1,0,0), \begin{pmatrix} -(\lambda_1 + \lambda_2 + \lambda_3) & \lambda_2 & \lambda_1 \\ 0 & -\gamma_t & 0 \\ 0 & 0 & -\gamma_k \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

In section 2.4 it will be seen that this representation of the bivariate Marshall-Olkin exponential distribution will make it possible to obtain a matrix representation for the bivariate $MMPP_2$.

2.2 The bivariate MMPP2

To formally define the bivariate $MMPP$ a two-state Markov process $J(t)$ with generator $\boldsymbol{Q}$ on $\{1,2\}$ is considered as for the $MMPP_2$. Whenever $J(t) = i$, it is said that the process is in state i and this status remains unchanged while $J(t)$ remains in this state. Specifically, the bivariate $MMPP_2$ behaves as follows: the initial state $i_0 \in \mathcal{S} = \{1,2\}$ is defined according to an initial probability vector $\boldsymbol{\alpha}= (\alpha_1, \alpha_2)$. The time during which the process remains in each state is given by two bivariate Marshall-Olkin exponential distribution. The first one models the sojourn times in state i. It is characterized by the parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, and the marginal distributions have mean $1/\gamma_{t1}$ and $1/\gamma_{k1}$ respectively, with

$$\gamma_{t1} = \lambda_1 + \lambda_3 \quad \text{and} \quad \gamma_{k1} = \lambda_2 + \lambda_3. \tag{4.7}$$

The second BVE associated with state $i = 2$ has parameters $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ and its marginal exponential distributions have mean $1/\gamma_{t2}$ and $1/\gamma_{k2}$ respectively, with

$$\gamma_{t2} = \omega_1 + \omega_3 \quad \text{and} \quad \gamma_{k2} = \omega_2 + \omega_3. \tag{4.8}$$

At the end of each bivariate Marshall-Olkin exponentially distributed sojourn time, two possible state transitions can occur. First, with probability a if $i = 1$ (b if $i = 2$), no failure occurs and the bivariate $MMPP_2$ enters into the other state $j \neq i$. Second, with probability $1 - a$ if $i = 1$ ($1 - b$ if $i = 2$), a failure is produced and the system continues in the same state.

This construction allows to create a bivariate sequence $\{(T_1, K_1), \ldots, (T_n, K_n)\}$. In relation with the train failure dataset, the times between failures will be denoted by T_i while the distances traveled K_i are recorded in the other variable.

On the other hand, it is noteworthy that the construction of the *MMPP* bivariate process facilitates the simulation of traces of this process as indicated in Algorithm 1 in Table 4.1. For a better understanding of the considered process, Figure 4.4 illustrates a realization of the bivariate $MMPP_2$, where (T_i, K_i) are represented in two "time" lines (one for times and the other one for distances). In this figure, it can be appreciated how both "time" lines have a common starting point ($t = 0$ and $k = 0$) and a common sequence of transitions through the states of the underlying process $\{s = 1, s = 2, s = 1, s = 1, \ldots\}$.

Summarizing, the bivariate $MMPP_2$ proposed in this paper is defined by two marginal processes with the same underlying process and two bivari-

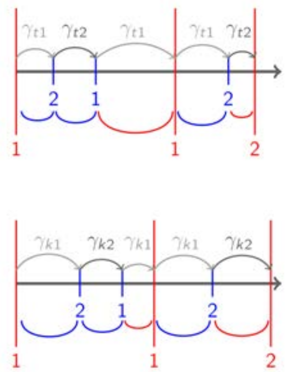


Figure 4.4: Transition diagram for two sequence of events generated by a bivariate $MMPP_2$, where the one above is for "time" and the one below is for "distances". The blue lines corresponds to transitions without failures and the red lines correspond to transitions with failures.

ate Marshall-Olkin exponential distributions. Essentially, the two marginal processes are running alongside, with the same transitions, the same underlying process and simultaneous failures, but each with different inter-failures "time" rates between failures. The structure of the marginal processes make it possible to capture the inter-dependence of the data. On the other hand, the existence of a single underlying process that relates to both marginal processes is what brings about the correlation between the two "timelines" (the intra-dependence in the data). Therefore, the bivariate $MMPP_2$ is fully described by eight parameters: the three parameters associated to each BVE $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ and $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ and the two probabilities associated to the Markov underlying process (a, b).

2.3 Identifiability of the bivariate MMPP$_2$

Identifiability problems occur when different representations of the process lead to the same likelihood functions for the observable data. It is well known that the MAP and $BMAP$ processes suffer from identifiability problem, but, on the other hand, Rydén (1996b) and Yera et al. (2019a) prove the

ALGORITHM 1

Algorithm to generate a trace from a bivariate $MMPP_2$.

0. Input: $\{a, b, \boldsymbol{\lambda}, \boldsymbol{\omega}\}$
1. Compute $\boldsymbol{\phi}$ as in (4.3).
2. Generate $p \sim U(0,1)$.
3. If $\phi_1 < p$, set $s_0 = 1$ else $s_0 = 2$.
4. Initialize $i = 1$, $j = 0$, $\boldsymbol{t} = \mathbf{0}$ and $\boldsymbol{k} = \mathbf{0}$.
5. While $i < n$ repeat:
   (a) Generate (X,Y) as a bivariate Marshall-Olkin exponential distribution (BVE).
   If $s_j = 1$, $(X, Y) \sim BVE(\boldsymbol{\lambda})$ else $(X, Y) \sim BVE(\boldsymbol{\omega})$
   (b) Set $(t_i, k_i) = (t_i, k_i) + (X, Y)$
   (c) Generate $p \sim U(0,1)$.
   (d) if ($s_j = 1$ and $p < a$) or ($s_j = 2$ and $p < b$), then (there is not failure) $s_{j+1} = s_j$. Else (there is a failure) $s_{j+1} = 2^{s_j} - 1$ and $i = i + 1$.
   (d) Set $j = j + 1$.
6. Output: $\{(t_1, k_1), \ldots, (t_n, k_n)\}$

Table 4.1: Algorithm to generate a trace of a bivariate $MMPP_2$

identifiability of the *MMPP* and *BMMPP*. In this section the identifiability of the $MMPP_2$ will be extended to the bivariate case. Let us first define the identifiability notion used in this paper.

Definition 11. *Let $B = \{\boldsymbol{\lambda}, \boldsymbol{\omega}, a, b\}$ be a representation of a bivariate $MMPP_2$ and let T_n and K_n denote the sequences between the $(n-1)$-th and the n-th event occurrence. Then B is said to identifiable if there exists* *no different parameterization* $\tilde{B} = \{\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\omega}}, \tilde{a}, \tilde{b}\}$ *such that*

$$\{(T_1, K_1), \ldots, (T_n, K_n)\} \stackrel{d}{=} \{(\tilde{T}_1, \tilde{K}_1), \ldots, (\tilde{T}_n, \tilde{K}_n)\} \quad \textit{for all } n \geq 1, \quad (4.9)$$

where $\tilde{T}_i$ *and* $\tilde{K}_i$ *are defined in analogous way as* T_i *and* K_i*, and where* $\stackrel{d}{=}$ *denotes equality in distribution.*

Theorem 4. *Let* $B = \{\boldsymbol{\lambda}, \boldsymbol{\omega}, a, b\}$ *and* $\tilde{B} = \{\tilde{\boldsymbol{\lambda}}, \tilde{\boldsymbol{\omega}}, \tilde{a}, \tilde{b}\}$ *be two different, but equivalent, representations of a bivariate* MMPP_2*. Then* $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}$*,* $\boldsymbol{\omega} = \tilde{\boldsymbol{\omega}}$ *and* $(a, b) = (\tilde{a}, \tilde{b})$*, except for a swap of* a *by* b *and* $\boldsymbol{\lambda}$ *by* $\boldsymbol{\omega}$*.*

See Appendix A for the proof.

2.4 Matrix representation

This section presents one of the most interesting findings regarding the bivariate $MMPP_2$. It is a matrix representation that facilitates obtaining analytical and algorithmic results in relation to the amounts of interest asociated with the process. Indeed, the bivariate $MMPP_2$ can be represented by $B = \{\boldsymbol{\phi}, \boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{R}\}$, where the initial probability vector is constructed from the stationary probability vector given in (4.3):

$$\boldsymbol{\phi} = (\phi_1, 0, 0, \phi_2, 0, 0)$$

Note that ϕ_1 and ϕ_2, defined in (4.3), only depend on the probabilities associated with the underlying process. The matrices governing transitions in which failures do not occur and occur are given by

$$\boldsymbol{D_0} = \left(\begin{array}{ccc|ccc} -(\lambda_1+\lambda_2+\lambda_3) & \lambda_2 & \lambda_1 & \lambda_3 a & 0 & 0 \\ 0 & -\gamma_{t1} & 0 & \gamma_{t1} a & 0 & 0 \\ 0 & 0 & -\gamma_{k1} & \gamma_{k1} a & 0 & 0 \\ \hline \omega_3 b & 0 & 0 & -(\omega_1+\omega_2+\omega_3) & \omega_2 & \omega_1 \\ \gamma_{t2} b & 0 & 0 & 0 & -\gamma_{t2} & 0 \\ \gamma_{k2} b & 0 & 0 & 0 & 0 & -\gamma_{k2} \end{array}\right),$$

and

$$\boldsymbol{D_1} = \left(\begin{array}{ccc|ccc} \lambda_3(1-a) & 0 & 0 & 0 & 0 & 0 \\ \gamma_{t1}(1-a) & 0 & 0 & 0 & 0 & 0 \\ \gamma_{k1}(1-a) & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \omega_3(1-b) & 0 & 0 \\ 0 & 0 & 0 & \gamma_{t2}(1-b) & 0 & 0 \\ 0 & 0 & 0 & \gamma_{k2}(1-b) & 0 & 0 \end{array}\right)$$

respectively. Finally

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Remark 2. *Note that the previous matrix representation is not a canonical* *representation of the process since it is possible to construct another matrix representation similar to this one that fully describes the process (see Appendix B for an equivalent matrix representation). However, the parameters used in the alternative matrix representation are the same than those used previously and described in Theorem 4. The in-depth study of the different matrix representations for this process and the search for a canonical representation is outside the scope of this work and will be addressed in the future.*

Under this representation it can be seen that the pair of variables (T_n, K_n) follows a multivariate phase type distribution MPH^* as defined in Kulkarni (1989) and Bladt and Nielsen (2017), with representation $\{\boldsymbol{\phi}, \boldsymbol{D_0}, \boldsymbol{R}\}$. This result will allow the use of Theorem 8.1.2. in Bladt and Nielsen (2017) in which the moment-generating function for an MHP^* is derived in order to obtain the moment-generating function until the occurrence of the first failure in the bivariate $MMPP_2$. We recall that $\Delta(a)$ denotes the diagonal matrix with vector a as diagonal.

Proposition 7. *Let $B = \{\boldsymbol{\phi}, \boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{R}\}$ be a representation of a bivariate $MMPP_2$ and let $\boldsymbol{A} = (T_1, K_1)$ be the records related to the first failure. Then there exists a $K > 0$ such that the moment-generating function for $\boldsymbol{A}$ (denoted by $H(\boldsymbol{\theta})$) exists and is given by*

$$H(\boldsymbol{\theta}) = E\left(e^{\boldsymbol{A}\boldsymbol{\theta}}\right) = \boldsymbol{\phi}\left(-\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta}) - \boldsymbol{D_0}\right)^{-1}\boldsymbol{D_1}\boldsymbol{e},$$

for any $s, t < K$ and with $\boldsymbol{\theta} = [\theta_1, \theta_2]^t$.

Proof. Due to $A \sim MPH^*(\boldsymbol{\phi}, \boldsymbol{D_0}, \boldsymbol{R})$, the Theorem 8.1.2. in Bladt and Nielsen (2017) can be applied directly to $\boldsymbol{A}$. Substituting $\boldsymbol{\alpha} = \boldsymbol{\phi}$, $\boldsymbol{S} = \boldsymbol{D_0}$

and $\boldsymbol{s} = \boldsymbol{D_1 e}$ the result is immediate. □

From Proposition 7, the formulae for moments and cross moments of the process can be derived.

Proposition 8. *Let $B = \{\boldsymbol{\phi}, \boldsymbol{D_0}, \boldsymbol{D_1}, \boldsymbol{R}\}$ be a representation of a bivariate $MMPP_2$, then the joint moments between $\{T\}_i$ and $\{K\}_i$ are given by*

$$\eta_{nm} = E\left(T_i^n K_i^m\right) = \boldsymbol{\phi} \sum_{i=1}^{(m+n)!} \left(\prod_{j=1}^{n+m} (-\boldsymbol{D_0})^{-1} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot \sigma_i(j)}) \right) \boldsymbol{e}, \tag{4.10}$$

where $\boldsymbol{R_{\cdot j}}$ is j^{th} column of $\boldsymbol{R}$, $\sigma_1, ..., \sigma_{(n+m)!}$ are the ordered permutations of duples of derivatives, and $\sigma_i(j) \in \{1, 2\}$ and is the i^{th} position of the permutation σ_i.

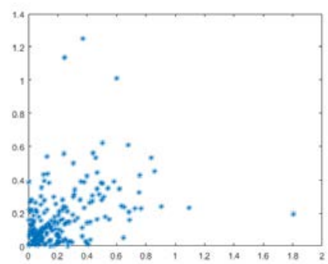
Proof. As the pair of the variables (T_i, K_i) follows a multivariate phase type distribution MPH^* Theorem 8.1.5 in Bladt and Nielsen (2017) can be applied directly to (T_i, K_i). Substituting $\boldsymbol{\alpha} = \boldsymbol{\phi}$, and $\boldsymbol{U} = \boldsymbol{(-D_0)}^{-1}$ the result is immediate. □

The joint moments that will be used in the estimation method are:

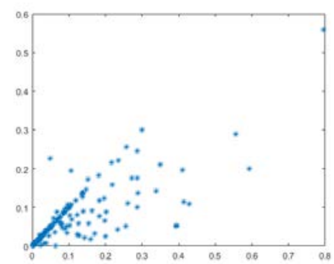
$$\begin{aligned}
\eta_{11} = E(TK) &= \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[\cdot 1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[\cdot 2]}} \qquad (4.11)\\
&\quad + \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[\cdot 2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[\cdot 1]}}\\
\eta_{21} = E(T^2K) &= 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.2]}}\\
&\quad + 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.1]}}\\
&\quad + 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.1]}}\\
\eta_{12} = E(TK^2) &= 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.1]}}\\
&\quad + 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.2]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.2]}}\\
&\quad + 2\boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-2}\boldsymbol{\Delta}(\boldsymbol{R_{[.1]}})(-\boldsymbol{D_0})^{-1}\boldsymbol{R_{[.2]}}
\end{aligned}$$

where $\boldsymbol{R}_{[.\boldsymbol{j}]}$ is the j^{th} column of $\boldsymbol{R}$.

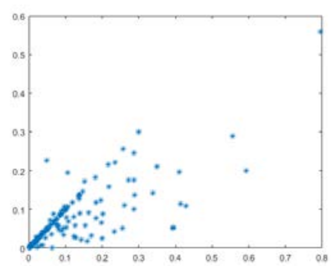
From the expressions previously obtained for η_{11}, η_{21} and η_{12} it is not difficult to deduce a closed expression for $corr(T, K)$. One of the advantages of the bivariate $MMPP_2$ that has been introduced in this work is the versatility with respect to the correlations that can be generated between the two sequences of events. Figure 4.5 shows this good property with four examples of scatter plots.



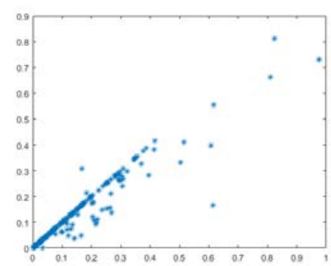
(a) $corr(T, K) = 0.38$



(b) $corr(T, K) = 0.64$



(c) $corr(T, K) = 0.84$



(d) $corr(T, K) = 0.94$

Figure 4.5: Scatter plot of simulated sequences generated from a bivariate $MMPP_2$.

In a similar way to the univariate case, from Proposition 7 the moment generating function is deduced for the sequences of the events.

Proposition 9. *The moment generating function of the n first consecutive* *events* $((T_1,K_1),(T_2,K_2),\dots,(T_n,K_n))$ *is given by*

$$
\begin{aligned}
f^*_{\{(T_1,K_1),\dots,(T_n,K_n)\}}(\boldsymbol{\theta_1},\dots,\boldsymbol{\theta_n}) &= E\left(e^{-\sum_{i=1}^{n}(T_i,K_i)\begin{pmatrix}\theta_{i1}\\ \theta_{i2}\end{pmatrix}} \right)\\
&= \boldsymbol{\phi}(-\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta}_1)-\boldsymbol{D}_0)^{-1}\boldsymbol{D}_1\times\dots\\
&\quad \times(-\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta}_n)-\boldsymbol{D}_0)^{-1}\boldsymbol{D}_1\boldsymbol{e}
\end{aligned}
$$

Corollary 1. *With* $X_1 = T_1$ *or* $X_1 = S_1$ *and* $Y_{1+n} = T_{1+n}$ *or* $Y_{1+n} = S_{1+n}$ *we have*

$$
E(X_1Y_{1+n}) = \boldsymbol{\phi}(-\boldsymbol{D_0})^{-1}\boldsymbol{R}_{[\cdot \boldsymbol{i}]}\boldsymbol{P}^{n-1}(-\boldsymbol{D_0})^{-1}\boldsymbol{R}_{[\cdot \boldsymbol{j}]}\boldsymbol{e}, \quad n = 1,2\dots
$$

where $\boldsymbol{P} = (-\boldsymbol{D_0})^{-1}\boldsymbol{D_1}$, $i = 1$ *for* $X_1 = T_1$, $i = 2$ *for* $X_1 = S_1$, $j = 1$ *for* $Y_{1+n} = T_{1+n}$, *and* $j = 2$ *for* $Y_{1+n} = S_{1+n}$.

Remark 3. *From Proposition 9 it can be verified that the marginal processes of the bivariate MMPP*$_2$ *defined in this paper, using the Marshall-Olkin bivariate exponential, are MMPP*$_2$ *with the following rate matrices:*

$$
B_t = \left\{ \boldsymbol{D_{0t}} = \begin{pmatrix} -\gamma_{t1} & \gamma_{t1}a \\ \gamma_{t2}b & -\gamma_{t2} \end{pmatrix}, \quad \boldsymbol{D_{1t}} = \begin{pmatrix} \gamma_{t1}(1-a) & 0 \\ 0 & \gamma_{t2}(1-b) \end{pmatrix} \right\} \tag{4.12}
$$

and

$$B_k = \left\{ \boldsymbol{D_{0k}} = \begin{pmatrix} -\gamma_{k1} & \gamma_{k1} a \\ \\ \gamma_{k2} b & -\gamma_{k2} \end{pmatrix}, \quad \boldsymbol{D_{1k}} = \begin{pmatrix} \gamma_{k1}(1-a) & 0 \\ \\ 0 & \gamma_{k2}(1-b) \end{pmatrix} \right\}. \tag{4.13}$$

Remark 4. *We have formulated Proposition 9 and Corollary 1 for the bivariate MMPP$_2$ process, but both holds in the greater generality of a sequence of correlated MPH^* variables. The only modification needed is to consider $\boldsymbol{\theta}$ vectors of higher dimension and replace T and X. Similarly, Corollary 1 can be extended to higher powers and to include more variables, but at the expense of loosing transparency.*

3 Inference for the bivariate *MMPP*$_2$

In this section, an approach for estimating the parameters of a bivariate *MMPP*$_2$ is proposed. Since, in practice the complete sequence of visited states of the underlying Markov process are not observed, the proposed procedure assumes that $\boldsymbol{t} = (t_1, t_2, ..., t_n)$, and $\boldsymbol{k} = (k_1, k_2, ..., k_n)$, are the only available information. An important issue to take into account is the complication when applying the likelihood principle to a singular measure. Therefore, the proposed fitting algorithm avoids evaluating or optimizing the likelihood. Indeed, the algorithm is a two-step approach. In the first one, the rate matrices of the marginal processes $\boldsymbol{D_{0t}}, \boldsymbol{D_{1t}}, \boldsymbol{D_{0k}}, \boldsymbol{D_{1k}}$ are estimated through a moments matching method defined by a standard optimization problem. The remaining parameters are estimated in a second step via an ABC algorithm. Section 3.2 illustrates the performance of the method on simulated data sets and finally, Section 3.3 addresses the modeling of the real train failures database described in Section 1.

3.1 The fitting algorithm

Bodrog et al. (2008) characterizes the $MMPP_2$ by the first three moments and first-lag auto-correlation coefficient of the inter-failure time distribution, $\mu_T(1), \mu_T(2), \mu_T(3)$ as in (4.4) and $\rho_T(1)$ as in (4.5). This implies that, for the bivariate $MMPP_2$ considered in this paper, the matrix representation of the marginal processes as in (4.12) and (4.13) will be characterized by a set of eight moments. Four of them are related to the variable (T), $\{\mu_T(1)$, $\mu_T(2)$, $\mu_T(3)$, $\rho_T(1)\}$ and the other four to the variable (K), $\{\mu_K(1)$, $\mu_K(2)$, $\mu_K(3)$, $\rho_K(1)\}$.

Carrizosa and Ramírez-Cobo (2014), using the results found by Bodrog et al. (2008), derive a moments matching method for estimating the parameters of a MAP_2, given a sequence of inter-failure times $\boldsymbol{t} = (t_1, t_2, ..., t_n)$. Posterior adaptations of this procedure can be found in Rodríguez et al. (2016a) and Yera et al. (2019a) to estimate the non-stationary MAP and $BMMPP_2$ respectively. In the case of the bivariate $MMPP_2$, it is an open problem to represent the process by a set of moments. However, a moments matching approach similar as in Carrizosa and Ramírez-Cobo (2014) can be designed to partially estimate its parameters. In particular, the parameters associated to the marginal processes $\boldsymbol{D_{0t}}$, $\boldsymbol{D_{1t}}$, $\boldsymbol{D_{0k}}$ and $\boldsymbol{D_{1k}}$ in terms of $\gamma_{t1}, \gamma_{t2}, \gamma_{k1}, \gamma_{k2}, a$ and b (see (4.12) and (4.13)), can be estimated as the solution of the following optimization problem:

$$(P0) \begin{cases} \min\limits_{\boldsymbol{\gamma},a,b} & \delta_0(a, b, \gamma_{t1}, \gamma_{t2}, \gamma_{k1}, \gamma_{k2}) \\ \text{s.t.} & \gamma_{t1}, \gamma_{t2}, \gamma_{k1}, \gamma_{k2} \geq 0, \\ & 0 \leq a, b \leq 1, \end{cases}$$

where the objective function is

$$\begin{aligned}\delta_0(\gamma_{t1},\gamma_{t2},\gamma_{k1},\gamma_{k2},a,b) &= [\rho_T(1)(a,b,\gamma_{t1},\gamma_{t2})-\bar{\rho}_T(1)]^2\\ &+[\rho_K(1)(a,b,\gamma_{k1},\gamma_{k2})-\bar{\rho}_K(1)]^2\\ &+\sum_{j=1}^3\left(\frac{\mu_T(j)(a,b,\gamma_{t1},\gamma_{t2})-\bar{\mu}_T(j)}{\bar{\mu}_T(j)}\right)^2\\ &+\sum_{j=1}^3\left(\frac{\mu_K(j)(a,b,\gamma_{k1},\gamma_{k2})-\bar{\mu}_K(j)}{\bar{\mu}_K(j)}\right)^2\end{aligned}$$

where $\bar{\mu}_T(i)$, for $i=1,2,3$ and $\bar{\rho}_T(1)$ denote the empirical moments associated to the first marginal process (computed from the sample $\boldsymbol{t}$), whereas $\bar{\mu}_K(i)$, for $i=1,2,3$ and $\bar{\rho}_K(1)$ denote the empirical moments associated to the second one (computed from the sample $\boldsymbol{k}$). Note that in the previous objective function, $\rho_T(1)$ and $\mu_T(i)$, for $i=1,2,3$ depend on a,b,γ_{t1} and γ_{t2}, while $\rho_K(1)$ and $\mu_K(i)$ for $i=1,2,3$ depend on a,b,γ_{k1} and γ_{k2}. Probabilities a and b are common for the two marginal process since both marginal $MMPP_2$ share the same underlying Markovian process. For more details of this procedure, see Step 1 of Algorithm 2. Note also that we propose to solve problem P0 a number I of times, so that the final solution will be the one that provides the lowest objective function. Each time, problem P0 is solved using a different starting point $(a(0),b(0),\gamma_{t1}(0),\gamma_{t2}(0),\gamma_{k1}(0),\gamma_{k2}(0))$ to avoid getting stuck at a poor local optimum. In practice, I is set to be equal to 100, which has proven in the numerical experiments to be high enough so as to get reasonable solutions. The final solution $\{\hat{a},\hat{b},\hat{\gamma}_{t1},\hat{\gamma}_{t2},\hat{\gamma}_{k1},\hat{\gamma}_{k2}\}$ shall be considered the starting point for the second step in the proposed fitting algorithm, an ABC approach whose output is the complete set of parameters characterizing the bivariate $MMPP_2$, that is $\{a,b,\boldsymbol{\lambda},\boldsymbol{\omega}\}$. From Step 1, the estimated values of a and b are already obtained With regards to the values of $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ note that since $\gamma_{t1}=\lambda_1+\lambda_3$, $\gamma_{t2}=\omega_1+\omega_3$, $\gamma_{k1}=\lambda_2+\lambda_3$ and $\gamma_{k2}=\omega_2+\omega_3$, see (4.7) and (4.8), the estimated values for $\lambda_1+\lambda_3$, $\lambda_2+\lambda_3$, $\omega_1+\omega_2$ and $\omega_2+\omega_3$ are obtained from the moment matching approach. On the other hand, as it has been commented previously, the likelihood function is not trivial to apply in the current setting with a singular measure.

Although simulated traces from the bivariate $MMPP_2$ are easy to generate (see Algorithm 1). In this setting, the ABC algorithm turns out suitable, see for example Csilléry et al. (2010), Marin et al. (2012) and Kypraios et al. (2017). The ABC algorithm is mathematically well-founded and applied in a wide variety of fields, but there are some issues that have to be carefully considered for a good performance, as the dimension of the parameter space. The larger the dimension of the parametric space, the more simulations are needed since, as Csilléry et al. (2010) points out, the probability of accepting the simulated values for the parameters under a given tolerance decreases exponentially with increasing dimensionality of the parameter space.

ALGORITHM 2

`Step 1`: `Moments matching approach`

0. `Input:` $\{\bar{\mu}_T(1), \bar{\mu}_T(2), \bar{\mu}_T(3), \bar{\rho}_T(1), \bar{\mu}_K(1), \bar{\mu}_K(1), \bar{\mu}_K(1), \bar{\rho}_K(1)\}$

1. `For` $i = 1, \ldots, I$ `repeat:`

   (a) `Randomly select a starting point` $\{a^{(i)}(0), b^{(i)}(0), \gamma_{t1}^{(i)}(0), \gamma_{t2}^{(i)}(0), \gamma_{k1}^{(i)}(0), \gamma_{k2}^{(i)}(0)\}$.

   (b) `Solve` $(P0)_i$ `and save the value of objective function` $\delta_0^{(i)}$ `and the solution` $\{\hat{a}^{(i)}, \hat{b}^{(i)}, \widehat{\gamma}_{t1}^{(i)}, \widehat{\gamma}_{k1}^{(i)}, \widehat{\gamma}_{t2}^{(i)}, \widehat{\gamma}_{k2}^{(i)}\}$.

2. `Obtain` $j = \arg\min_i \delta_0^{(i)}$ `and set` $\delta_0 = \delta_0^{(j)}, \widehat{\gamma}_{t1} = \widehat{\gamma}_{t1}^{(j)}$, $\widehat{\gamma}_{t2} = \widehat{\gamma}_{t2}^{(j)}, \widehat{\gamma}_{k1} = \widehat{\gamma}_{k1}^{(j)}, \widehat{\gamma}_{k2} = \widehat{\gamma}_{k2}^{(j)}, \hat{a} = \hat{a}^{(j)}$ `and` $\hat{b} = \hat{b}^{(j)}$

3. `Output:` $\{\hat{a}, \hat{b}, \widehat{\gamma}_{t1}, \widehat{\gamma}_{t2}, \widehat{\gamma}_{k1}, \widehat{\gamma}_{k2}\}$

Step 1 in the fitting approach: moments matching method to estimate the parameters of the marginal components in the bivariate $MMPP_2$

In the case of the bivariate $MMPP_2$, estimates for a, b, $\lambda_1 + \lambda_3$ (γ_{t1}), $\lambda_2 + \lambda_3$ (γ_{k1}), $\omega_1 + \omega_3$ (γ_{t2}) and $\omega_2 + \omega_3$ (γ_{k2}) are obtained in Step 1. Taking into account expressions (4.7) and (4.8) and setting prior distributions for λ_3 and ω_3 as

$$\lambda_3 \sim Unif(0, \min[\widehat{\gamma}_{t1}, \widehat{\gamma}_{k1}]), \quad \omega_3 \sim Unif(0, \min[\widehat{\gamma}_{t2}, \widehat{\gamma}_{k2}]), \tag{4.14}$$

then, a simple ABC algorithm in terms of only two parameters can be easily formulated. At each iteration $i \in 1, \ldots, I$, values of λ_3 and ω_3 are generated from the prior $\pi(\cdot)$ as in (4.14). Then, the values of λ_1, λ_2, ω_1 and ω_2 are obtained according to (4.7) and (4.8). Then, a sample $\boldsymbol{s}^{(i)} = \{(t_1, k_1), \ldots, (t_n, k_n)\}$ from a bivariate $MMPP_2$ with parameters $(\hat{a}, \hat{b}, \boldsymbol{\lambda}^{(i)}, \boldsymbol{\omega}^{(i)})$ is simulated according to Algorithm 1. If the generated sample $\boldsymbol{s}^{(i)}$ is too different from the observed data $\boldsymbol{s} = \{(t_1, k_1), \ldots, (t_n, k_n)\}$, the set of parameter is discarded. For this purpose a distance measure and a tolerance, $\epsilon > 0$, are usually established. The level of discrepancy between the generated sample at iteration i and the original sample shall be measured according to:

$$\delta_1(s^{(i)}, s) = \sum_{l=1, j=1}^{2} \left(\frac{\bar{\eta}_{lj}(s^{(i)}) - \bar{\eta}_{lj}(s)}{\bar{\eta}_{lj}(s)} \right)^2, \tag{4.15}$$

where $\bar{\eta}_{lj}$, for $l, j = 1, 2$ denote the first empirical joint moments associated to the bivariate $MMPP_2$ process $(E(TK), E(T^2K), E(TK^2))$ as in (4.11). Expressions similar as (4.15), in which with a structure similar to this one, theoretical and estimated moments intervene, have already been used as an objective function in problems of adjustment of moments for *BMAP*s processes (see for example Rodríguez et al. (2015) and Yera et al. (2019b)). The rationale behind this selection is the fact that for all iterations, the simulated samples come from a bivariate $MMPP_2$ processes with the same marginal moments (a consequence of constant $\hat{\gamma}_{t1}, \hat{\gamma}_{t2}, \hat{\gamma}_{k1}, \hat{\gamma}_{t1}, \hat{a}$ and $\hat{b}$). Therefore, it makes sense to include joint moments in the distance measure. From extensive, empirical experiments it has been observed that η_{11}, η_{12} and η_{21} in combination with the eight marginal moments used in Step 1 $(\mu_T(1), \mu_T(2), \mu_T(3), \rho_T(1), \mu_K(1), \mu_K(2), \mu_K(3)\rho_K(1))$ are enough to characterize the parameters of the bivariate process. With respect to the tolerance level, instead of fixing a specific value of epsilon, we proceed in analogous way by keeping the 1% of the samples with smallest differences from the original sample. Finally, the estimated parameters are average values among the selected proportion. For a summary of the ABC procedure, see Step 2 of Algorithm 2.

ALGORITHM 2

`Step 2:` `ABC approach`

0. `Input:` $\{\bar{\eta}_{11}, \bar{\eta}_{12}, \bar{\eta}_{21}, \hat{a}, \hat{b}, \widehat{\gamma}_{t1}, \widehat{\gamma}_{k1}, \widehat{\gamma}_{t2}, \widehat{\gamma}_{k2}\}$

1. `For` $i = 1, \ldots, I_2$ `do repeat:`

   (a) `Generate` $(\lambda_3^{(i)}, \omega_3^{(i)})$ `from the prior distribution` $\pi(\cdot)$

   (b) `Obtain`

$$\begin{aligned} \lambda_1^{(i)} &= \hat{\gamma}_{t1} - \lambda_3^{(i)} \\ \lambda_2^{(i)} &= \hat{\gamma}_{k1} - \lambda_3^{(i)} \\ \omega_2^{(i)} &= \hat{\gamma}_{t2} - \omega_3^{(i)} \\ \omega_2^{(i)} &= \hat{\gamma}_{k2} - \omega_3^{(i)} \end{aligned}$$

   (c) `Simulate a sample` $\boldsymbol{s}^{(i)}$ `from the likelihood` $f(\cdot \mid \hat{a}, \hat{b}, \boldsymbol{\lambda}^{(i)}, \boldsymbol{\omega}^{(i)})$.

   (c) `Compute the moments` $\bar{\eta}_{11}, \bar{\eta}_{12}, \bar{\eta}_{21}$ `associated to the generated sample` $\boldsymbol{s}^{(i)}$.

   (d) `Compute` $\delta_1^{(i)}(\boldsymbol{s}^{(i)}, \boldsymbol{s})$ `as in (4.15)`

4. `The 1% of the sampled values with the smallest differences from the real data are accepted.`

5. `The Bayesian estimates is computed as the average of the accepted values.`

6. `Output:` $\{\hat{a}, \hat{b}, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3\}$

Step 2 in the fitting approach: ABC method to estimate the parameters of the bivariate $MMPP_2$

3.2 A simulation study

The aim of this section is to illustrate the behavior of the procedure described in Section 3.1 on the basis of two simulated data sets. Each simulated data set consists of a sequence of 1000 pairs of events $(\boldsymbol{t}, \boldsymbol{k}) = \{(t_1, k_1), (t_2, k_2), ..., (t_n, k_n)\}$ simulated from two different bivariate $MMPP_2$s, whose parameter sets $\{a, b, \boldsymbol{\lambda}, \boldsymbol{\omega}\}$ are listed in Table 4.2 *Generator Process* columns.

The first example considers a simulated sample from a bivariate $MMPP_2$ with low intra-dependence (both marginal processes present a the first lag autocorrelation coefficient around 0.2) and very high inter-dependence (correlation between times and distances around 0.9). The second considered trace is from a bivariate $MMPP_2$ with relatively high intra-dependence (with an autocorrelation for both marginal processes of around 0.4) and a moderate intra-dependence (correlation between times and distances around 0.7). The results obtained after running the fitting approach are shown in Table 4.2.

The first eight rows in Table 4.2 show the parameters of the process (the real ones under the column *Generator Process* and the estimated ones under the column *Estimated*). The ninth to sixteenth rows depict the marginal empirical moments that characterize marginal processes (theoretical and empirical in parenthesis in the *Generator Process* column and estimated ones in the *Estimation* column). The analogous can be found from row 17th to row 19th, but regarding the joint moments. The penultimate row shows the correlation between inter-failures times and distances, and finally, the last row shows the running time (measured in seconds) in an Intel Xeon Six-cores 3.6 GHz with 12 threads processor with 128Gb of memory ram (for a prototype code written in MATLAB©).

Some comments arise from the results presented in Table 4.2. On one hand, it should be pointed out the good performance of the method for estimating the moments of the bivariate process (both marginal and joint) and in particular, the correlation between the inter-failure times and distances. Parameters a and b are also well fitted, a fact that also happens for λ_3 and ω_3 since the ABC approach is specifically designed in terms of them. However, it has been observed that estimations for the rest of parameters $(\lambda_1, \lambda_2, \omega_1$ and $\omega_2)$ may be less good for some cases. Consider for example, the first simulated trace. From Step 1 in the fitting approach, the values

	Example 1		Example 2	
	Generator Process	*Estimation*	*Generator Process*	*Estimation*
a	0.02	0.02	0.008	0.008
b	0.44	0.44	0.08	0.09
λ_1	0.82	0.68	4.11	4.28
λ_2	0.40	0.31	1.79	1.62
λ_3	1.86	1.84	5.95	5.66
ω_1	2.35×10^{-2}	2.51×10^{-2}	0.12	0.12
ω_2	5.27×10^{-3}	7.91×10^{-3}	0.12	0.12
ω_3	0.24	0.22	0.33	0.33
$\mu_T(1)$	0.58 (0.57)	0.57	0.29 (0.28)	0.28
$\mu_T(2)$	1.99 (1.92)	1.92	0.88 (0.87)	0.87
$\mu_T(3)$	20.16 (20.40)	20.42	5.67 (5.63)	5.65
$\rho_T(1)$	0.22 (0.21)	0.22	0.41 (0.40)	0.40
$\mu_K(1)$	0.66 (0.66)	0.65	0.31 (0.32)	0.32
$\mu_K(2)$	2.38 (2.30)	2.31	0.89 (0.89)	0.90
$\mu_T(3)$	25.37 (25.78)	25.75	5.72 (5.79)	5.77
$\rho_K(1)$	0.21 (0.22)	0.21	0.39 (0.40)	0.39
η_{11}	2.02 (1.99)	1.93	0.69 (0.70)	0.69
η_{21}	20.10 (19.26)	19.98	3.84 (3.90)	3.85
η_{12}	21.27 (21.87)	21.16	3.85 (3.81)	3.88
$Cor(T, K)$	0.91 (0.90)	0.90	0.76 (0.75)	0.76
running time	-	219.79	-	213.89

Table 4.2: Comparison between the theoretical, empirical (within parentheses) and estimated values for obtained with Algorithm 1 for two examples.

of $\gamma_{t1} = \lambda_1 + \lambda_3$ and $\gamma_{k1} = \lambda_2 + \lambda_3$ (equal to 2.53 and 2.26) are estimated as 2.53 and 2.25. Since the generated values of λ_3 in Step 2 (ABC) are upper-bounded by $\min\{\hat{\gamma}_{t1}, \hat{\gamma}_{k1}\}$, then the results under the ABC approach will be better or worse depending on the estimates for γ_{t1} and γ_{k1}. In the previous example, the final estimation for λ_2 turns out slightly better than that of λ_1, as expected. A similar fact occurs in the case of the parameter ω_3.

An additional comment regards the computational time of the proposed fitting approach. For the considered examples, the total running time is around 6 minutes, where the most of computational cost is due to the ABC algorithm. The matching moment approach defining Step 1 is not expensive from a computational point of view (see Yera et al. (2019b)). However, it is known from the literature that the ABC algorithm is time consuming, in particular for high values of I_2 (number of iterations or number of times where traces are simulated), see Minter and Retkute (2019). In our case, such value, I_2, is set as 10000 which provides a good compromise between performance of the inference approach and computing time.

In order to explore in more depth the results under the ABC algorithm, consider Figure 4.6 and Table 4.2. Figure 4.6 shows the evolution of the estimation of parameter λ_3 as the acceptance percentage considered in the ABC Algorithm varies. Note that in left panel (Example 1) the estimated value with 10% acceptance is 1.14, which is very close to the mean value of the prior distribution of $\hat{\lambda}_3$ $(\min(\hat{\gamma}_{t1}, \hat{\gamma}_{k1})/2)$. As the acceptance rate decreases, $\hat{\lambda}_3$ approaches to λ_3, as expected. A similar behavior can be seen in the right panel (Example 2). Table 4.3 shows the whole set of parameters $(\boldsymbol{\lambda}, \boldsymbol{\omega})$ for different acceptance rates. Again, it can be seen that as the acceptance percentage decreases, the estimate becomes more accurate.

3.3 Case study: failures in train

In this section the performance of the approach for fitting the real database concerning failures in trains and described in Section 1 is illustrated. In particular, the fitting approach described in Section 3.1 shall be applied to

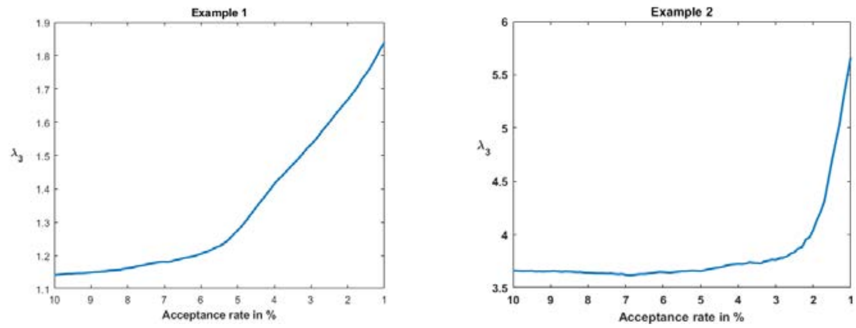


Figure 4.6: Evolution of the estimated value λ_3 for different acceptance percentages in the ABC algorithm (Step 2 in the fitting approach). The left (right) panel refers to the first (second) simulated example.

model failures in the trains numbered 35 and 36. A detailed exploratory analysis of the data is made in Pievatolo and Ruggeri (2010). Train 35 went into operation on 20/12/90 and in the almost 8 years, it traveled 359908 km and suffered 47 failures. The average, median, variation coefficient, minimum and maximum value of the inter-failures times for train 35 are 71.12, 24, 1.92, 1 and 736 days, respectively, which suggests a right-skewed distribution with a tail longer than that of an exponential distribution, a fact also deduced from Figure 1.5. Something similar occurs with the inter-failures distances, with the average, median, variation coefficient, minimum and maximum equal to 8778, 3592, 1.6842, 87 and 75998 kilometers. On the other hand, for train 36 there are records between 04/09/90 and 04/09/98, period in which it covered 379709 km and whose doors failed 51 times. The analysis carried out on the non-exponentiality of the inter-failures times and distances for train 35 applies also for train 36. Empirically, the consecutive inter-failure times and distances are not independent since the first lag autocorrelation coefficients are equal to 0.13 and 0.21, for trains 35 and 36, respectively. There is also interdependence among the traces, reflected by an empirical correlation coefficient equal to 0.97 and 0.93 for trains 35 and 36, respectively. The non-exponentiality of the traces in combination with both intra and inter-dependence makes the bivariate $MMPP_2$ proposed in

	Example 1					
	λ_1	λ_2	λ_3	ω_1	ω_2	ω_3
Generating process	0.82	0.40	1.86	0.02	0.005	0.24
1%	0.69	0.32	1.84	0.03	0.008	0.22
5%	1.25	0.88	1.27	0.02	0.007	0.22
10%	1.39	1.01	1.14	0.03	0.01	0.21
	Example 2					
	λ_1	λ_2	λ_3	ω_1	ω_2	ω_3
Generating process	4.11	1.79	5.95	0.12	0.12	0.33
1%	4.28	1.62	5.66	0.12	0.12	0.33
5%	6.29	3.63	3.66	0.12	0.12	0.33
10%	6.28	3.63	3.66	0.12	0.12	0.33

Table 4.3: Estimates for $\boldsymbol{\lambda}$ and $\boldsymbol{\omega}$ when varying the acceptance percentage in the ABC algorithm (Step 2 in the fitting approach)

this paper a suitable model for fitting the data.

The fitting approach described in Section 3.1 is run on the real traces to obtain estimates for the empirical quantities. The results are shown in Table 4.4, where the empirical marginal characterizing moments (first to eighth row), joint moments (ninth to eleventh rows) and correlation coefficient (twelfth row) are shown as well as their estimates.

The good performance of the fitted models is also supported by Figures 4.7 and 4.8. In particular, Figure 4.7 shows the fit to the empirical distribution function of the inter-failure times and inter-failure distances, see expression (4.2). On the other hand, Figure 4.8 concerns the estimation of conditional joint probabilities $P(T \mid K)$, for which a close-form expression is unknown, up to the authors. Once the estimated model is obtained, then, traces of the same size as the real ones ($n = 47$ for train 35 and $n = 51$ for train 36) are simulated 1000 times. Estimations of $P(T < t \mid K < k)$ are average values of the sampled frequencies. The left panel of Figure 4.8 concerns train 35, where the probability of having a failure in less than approximately six months given that a failure was observed in less than $k \in [0, 10000]$ km is estimated. The analogous probability is estimated in the right panel, for train 36 and in a period less than 3 months.

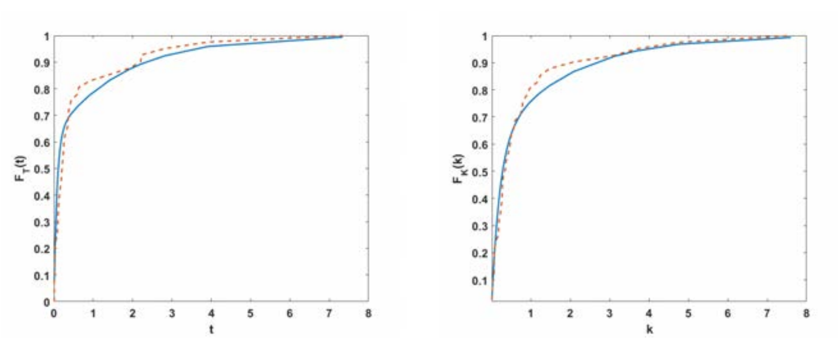


Figure 4.7: Estimated cdf (dashed line) under the bivariate $MMPP_2$ versus the empirical cdf (solid line) of the inter-failures times (left panel) and distances (right panel) for train 35.

Figure 4.9 shows the estimated expected number of failures occurred in

	Train 35		Train 36	
	Emp	Est biv *MMPP*	Emp	Est biv *MMPP*
$\mu_T(1)$	71.12	71.07	64.73	63.83
$\mu_T(2)$	2.33×10^4	2.35×10^4	1.11×10^4	1.07×10^4
$\mu_T(3)$	1.25×10^7	1.24×10^7	2.64×10^6	2.71×10^6
$\rho_T(1)$	0.13	0.13	0.21	0.19
$\mu_K(1)$	8.78×10^3	8.79×10^3	8.44×10^3	8.23×10^3
$\mu_K(2)$	2.90×10^8	2.88×10^8	1.83×10^8	1.79×10^8
$\mu_K(1)$	1.58×10^{13}	1.59×10^{13}	5.68×10^{12}	5.81×10^{12}
$\rho_K(1)$	0.12	0.12	0.26	0.20
η_{11}	2.54×10^6	2.52×10^6	1.36×10^6	1.22×10^6
η_{21}	1.33×10^6	1.30×10^6	3.22×10^8	2.95×10^8
η_{12}	1.44×10^{11}	1.40×10^{11}	4.18×10^{10}	3.94×10^{10}
$Corr(T, K)$	0.97	0.96	0.93	0.82

Table 4.4: Empirical and estimated moments by the bivariate $MMPP_2$ for both trains 35 and 36.

the time recorded for trains 35 and 36. It is possible to observe that train 35 is subject to an abrupt increased failure rate after the initial period and then heavy intervention likely occurred since the failure rate went down. Train 36 is subject to an increase in the failure rate which then decreases (both changes occur over long periods of time) when better reliability begins to be appreciated. Finally, Figure 4.10 shows the estimation for train 36 of the expected number of failures in different joint time and distance intervals.

4 Conclusions

A bivariate extension of the two-state Markov modulated Poisson process is considered in this paper. This process allows for the modeling of non-

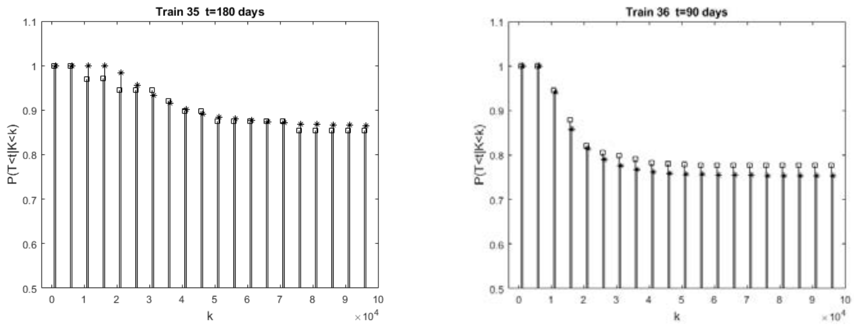


Figure 4.8: Comparison between the estimated (asterisk) and empirical (square) values of the conditional probabilities $p(T < t|K < k)$ for train 35 (left panel) and 36 (right panel).

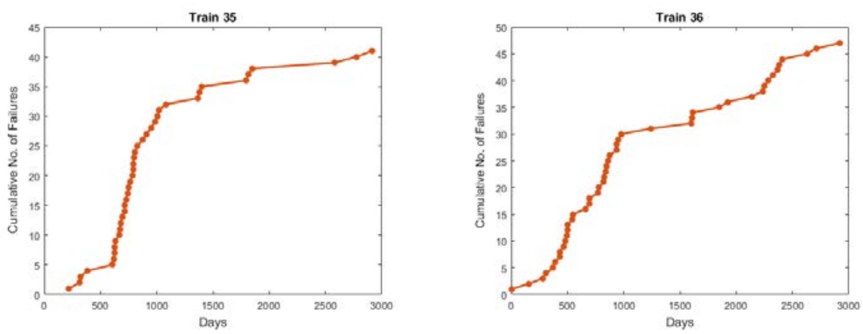


Figure 4.9: Estimated mean of number of failures under the bivariate $MMPP_2$ for train 35 (left panel) and train 36 (right panel).

exponential bivariate traces presenting inter and intra-dependence, properties that make the model suitable either in reliability or other real contexts. Some properties concerning the novel model are shown, in particular its identifiability, inherited from the marginal processes, and crucial if inference is to be undertaken.

Once the process is properly described, a fitting approach is presented. The method combines a matching moment approach (Step 1) with an ABC algorithm (Step 2). The first step helps alleviating the computational cost

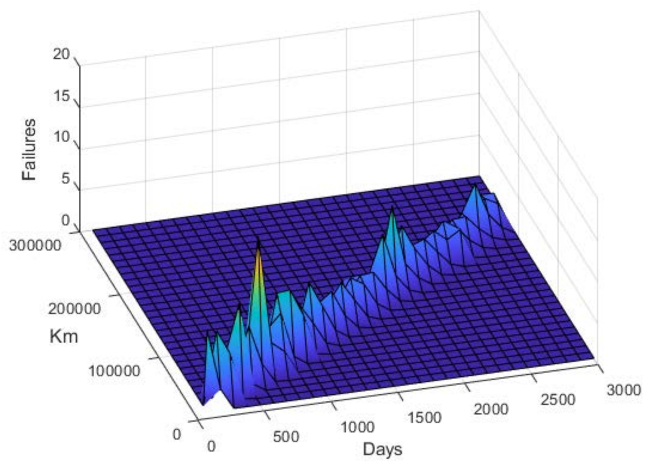


Figure 4.10: Estimation of the expected number of failures under the bivariate $MMPP_2$ for train 36.

inherent in the ABC since the number of parameters to be estimated go from 8 to 2.

The methodology is illustrated for both simulated and real datasets. In particular, an application to real failures data set concerning train failures is presented. The results show the potential of the proposed bivariate model as well as the right performance of the fitting method.

Prospects regarding this work concern both theoretical and applied issues. Some theoretical problems to be considered are as follows. First, find closed expressions for quantities of interest as joint probabilities in terms of (T) and (K) or the joint predictive distributions. Also, it is of interest to obtain probabilities of the counting process $N_T(t)$ $(N_K(k))$, number of events up to $t(k)$, as in Neuts and Li (1997). A third theoretical problem is the extension of the process proposed in this paper to its batch counterpart, where failures can occur in simultaneous way (see Yera et al. (2019a,b)) for results concerning the batch counterpart of the $MMPP_2$). Also obtaining some set of moments that characterize the bivariate $MMPP_2$ would be very useful for the estimation of the process. Finally obtaining a canonical matrix representation would allow a more in-depth study of some theoretical properties of this process.

From an applied viewpoint, it is of interest to develop a more sophisticated version of the ABC algorithm, so that a low number of parameters is sampled, but where the existing dependence between the estimates from Step 1 and Step 2 is mitigated. It would be important also to estimate how many failures will occur in future time and km intervals, both in terms of expected number and probability of no failure in a given [time, km] interval. Work on these issues is underway.

Acknowledgments

Research partially supported by research grants and projects MTM2015-65915-R and ECO2015-66593-P (Ministerio de Economía y Competitividad, Spain), P11-FQM-7603, FQM-329 (Junta de Andalucía, Spain) and Fundación BBVA and Comunidad de Madrid through a Catedra de Excelencia.

Appendix A: Proof of Theorem 4

It is clear that if the representations B and $\bar{B}$ are equivalent, their respective marginal $MMPP_2$ will also be equivalent. Since Rydén (1996b) proves that the *MMPP* is identifiable except by permutations of states, the probabilities associated to the underlying process for the two marginals are the same (except by permutation); that is,

$$a = \tilde{a}, \quad b = \tilde{b}, \tag{4.16}$$

and the failure occurrence rates of the marginal processes also coincide, (except by permutation of the vector $\boldsymbol{\lambda}$ by the vector $\boldsymbol{\omega}$), which implies that,

$$\lambda_j + \lambda_3 = \tilde{\lambda}_j + \tilde{\lambda}_3, \quad \omega_j + \omega_3 = \tilde{\omega}_j + \tilde{\omega}_3, \quad \text{for } j = \{1, 2\}. \tag{4.17}$$

On the other hand, from (4.9) it can be proven that the bivariate exponential distribution associated to B and $\tilde{B}$ are equally distributed. Note that taking $n = 1$ in (4.9), the following equality is obtained:

$$(T_1, K_1) \stackrel{d}{=} (\tilde{T}_1, \tilde{K}_1). \tag{4.18}$$

(T_1, K_1) can be rewritten as the sum of N bivariate exponential distribution, where $N - 1$ is the number of times the underlying process changes state before the first failure occurs. On the other hand, since the underlying process is the same for both processes (4.18) is equivalent to

$$(X_1 + \cdots + X_N, Y_1 + \cdots + Y_N) \stackrel{d}{=} (\tilde{X}_1 + \cdots + \tilde{X}_N, \tilde{Y}_1 + \cdots + \tilde{Y}_N), \tag{4.19}$$

and it is possible to take conditional distribution on the initial state and the number of changes state of the underlying process in (4.19) obtaining

$$(X_1+\cdots+X_N, Y_1+\cdots+Y_N|N = 1, s_0 = i) \stackrel{d}{=} (\tilde{X}_1+\cdots+\tilde{X}_N, \tilde{Y}_1+\cdots+\tilde{Y}_N|N = 1, s_0 = i).$$

Hence the bivariate exponential distributions associated to B and $\tilde{B}$ have to be equally distributed and consequently they have the same moments. Therefore, using (4.17) and (4.6), the following relationship is obtained

$$\lambda_1 + \lambda_2 + \lambda_3 = \tilde{\lambda}_1 + \tilde{\lambda}_2 + \tilde{\lambda}_3, \quad \omega_1 + \omega_2 + \omega_3 = \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3. \tag{4.20}$$

Then, from (4.17) and (4.20), $\lambda_i = \tilde{\lambda}_i$ and $\omega_i = \tilde{\omega}_i$, for $i = \{1 \ldots 3\}$, which implies the identifiability of the process in the terms defined in the paper.

Equivalent Matrix Representation

An alternative matrix representation for the bivariate *MMPP* is given by the initial probability vector (4.3):

$$\boldsymbol{\phi} = \left(\phi_1 \frac{\lambda_2}{\lambda}, \phi_1 \frac{\lambda_1}{\lambda}, \phi_1 \frac{\lambda_3}{\lambda}, \phi_2 \frac{\omega_2}{\omega}, \phi_2 \frac{\omega_1}{\omega}, \phi_2 \frac{\omega_3}{\omega} \right),$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ and $\omega = \omega_1 + \omega_2 + \omega_3$, and the matrices

$$\boldsymbol{D_0} = \left(\begin{array}{ccc|ccc} -\gamma_{t1} & 0 & \gamma_{t1} & 0 & 0 & 0 \\ 0 & -\gamma_{k1} & \gamma_{k1} & 0 & 0 & 0 \\ 0 & 0 & -\lambda & \frac{\lambda}{\omega}\omega_2 a & \frac{\lambda}{\omega}\omega_1 a & \frac{\lambda}{\omega}\omega_3 a \\ \hline 0b & 0 & 0 & -\gamma_{t2} & 0 & \gamma_{t2} \\ 0 & 0 & 0 & 0 & -\gamma_{k2} & \gamma_{k2} \\ \frac{\omega}{\lambda}\lambda_2 b & \frac{\omega}{\lambda}\lambda_1 b & \frac{\omega}{\lambda}\lambda_3 b & 0 & 0 & -(\omega) \end{array}\right),$$

$$\boldsymbol{D_1} = \left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ (1-a)\lambda_2 & (1-a)\lambda_1 & (1-a)\lambda_3 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_2(1-b) & \omega_1(1-b) & \omega_3(1-b) \end{array}\right)$$

and

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Note that the parameters associated to this alternative representation are the same than for the previous one but they offer a different arrangement in the matrix representation.

References

Al-Garni, A. Z., Sahin, A. Z., Al-Ghamdi, A. S., and Al-Kaabi, S. A. (1999). Reliability analysis of aeroplane brakes. *Quality and reliability engineering international*, 15(2):143–150.

Assaf, D., Langberg, N. A., Savits, T. H., and Shaked, M. (1984). Multivariate phase-type distributions. *Operations Research*, 32(3):688–702.

Badescu, A. L., Cheung, E. C., and Landriault, D. (2009). Dependent risk models with bivariate phase-type distributions. *Journal of Applied Probability*, 46(1):113–131.

Bearfield, G. and Marsh, W. (2005). Generalising event trees using Bayesian networks with a case study of train derailment. In *International Conference on Computer Safety, Reliability, and Security*, pages 52–66. Springer.

Bladt, M. and Nielsen, B. F. (2017). *Matrix-exponential distributions in applied Probability*, volume 81. Springer.

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.

Buchholz, P. (2003). An EM-algorithm for MAP fitting from real traffic data. In *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 218–236. Springer.

Cai, J. and Li, H. (2005). Multivariate risk model of phase type. *Insurance: Mathematics and Economics*, 36(2):137–152.

Carrizosa, E. and Ramírez-Cobo, P. (2014). Maximum likelihood estimation in the two-state Markovian arrival process. *arXiv preprint arXiv:1401.3105*, Working paper.

Chakravarthy, S. R. (2010). Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science.*

Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution*, 25(7):410–418.

Decò, A., Frangopol, D. M., and Zhu, B. (2012). Reliability and redundancy assessment of ships under different operational conditions. *Engineering Structures*, 42:457–471.

Eliashberg, J., Singpurwalla, N. D., and Wilson, S. P. (1997). Calculating the reserve for a time and usage indexed warranty. *Management Science*, 43(7):966–975.

Fang, C. and Das, P. K. (2005). Survivability and reliability of damaged ships after collision and grounding. *Ocean Engineering*, 32(3-4):293–307.

Fischer, W. and Meier-Hellstern, K. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Performance evaluation*, 18(2):149–171.

Goel, L. and Gupta, P. (1984). Analysis of a two-engine aeroplane model with two types of failure and preventive maintenance. *Microelectronics Reliability*, 24(4):663–666.

Griffiths, R., Milne, R., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21(3):238–255.

Griffiths, R. C. and Milne, R. K. (1978). A class of bivariate poisson processes. *Journal of Multivariate Analysis*, 8(3):380–395.

Ivanov, L. D. (2009). Challenges and possible solutions of the time-variant reliability of ship's hull girder. *Ships and Offshore Structures*, 4(3):215–228.

Jafarian, E. and Rezvani, M. (2012). Application of fuzzy fault tree analysis for evaluation of railway safety risks: an evaluation of root causes for passenger train derailment. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226(1):14–25.

Karim, M. R. (2008). Modelling sales lag and reliability of an automobile component from warranty database. *International Journal of Reliability and Safety*, 2(3):234–247.

Kızılaslan, F. and Nadar, M. (2018). Estimation of reliability in a multicomponent stress–strength model based on a bivariate Kumaraswamy distribution. *Statistical Papers*, 59(1):307–340.

Klemm, A., Lindemann, C., and Lohmann, M. (2003). Modeling IP traffic using Batch Markovian Arrival Process. *Performance Evaluation*, 54(2):149–173.

Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operations research*, 37(1):151–158.

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical biosciences*, 287:42–53.

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian Computational methods. *Statistics and Computing*, 22(6):1167–1180.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Minter, A. and Retkute, R. (2019). Approximate Bayesian Computation for infectious disease modelling. *Epidemics*, page 100368.

Nadar, M. and Kızılaslan, F. (2015). Estimation of reliability in a multicomponent stress-strength model based on a Marshall-Olkin bivariate Weibull distribution. *IEEE Transactions on Reliability*, 65(1):370–380.

Navas, M. A., Sancho, C., and Carpio, J. (2017). Reliability analysis in railway repairable systems. *International Journal of Quality & Reliability Management*, 34(8):1373–1398.

Neuts, M. and Li, J. (1997). *An algorithm for the $P(n,t)$ matrices of a continuous BMAP*, volume 183 of *Lectures notes in Pure and Applied Mathematics*, pages 7–19. Srinivas R. Chakravarthy and Attahiru, S. Alfa, editors. NY: Marcel Dekker, Inc.

Pievatolo, A. and Ruggeri, F. (2010). Bayesian modelling of train door reliability. *The Oxford Handbook of Applied Bayesian Analysis. Oxford University Press, Oxford*, pages 271–294.

Pievatolo, A., Ruggeri, F., and Argiento, R. (2003). Bayesian analysis and prediction of failures in underground trains. *Quality and Reliability Engineering International*, 19(4):327–336.

Ramírez-Cobo, P. and Carrizosa, E. (2012). A note on the dependence structure of the two-state Markovian arrival process. *Journal of Applied Probability*, 49:295–302.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2008). Bayesian analysis of a queueing system with a long-tailed arrival process. *Communications in StatisticsSimulation and Computation*, 37(4):697–712.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010). Nonidentifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47(3):630–649.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2017). Bayesian analysis of the stationary MAP_2. *Bayesian Analysis*, 12(4):1163–1194.

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 and System Safety*, 134:126–133.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016a). Dependence patterns for modeling simultaneous events. *Reliability Engineering and System Safety*, 154:19–30.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016b). Nonidentifiability of the two-state $BMAP$. *Methodology and Computing in Applied Probability*, 18(1):81–106.

Ruggeri, F. (2006). On the reliability of repairable systems: methods and applications. In *Progress in Industrial Mathematics at ECMI 2004*, pages 535–553. Springer.

Rydén, T. (1996a). An EM algorithm for estimation in Markov modulated Poisson processes. *Computational Statistics & Data Analysis*, 21(4):431–447.

Rydén, T. (1996b). 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.

Sheikh, A. K., Al-Garni, A. Z., and Affan Badar, M. (1996). Reliability analysis of aeroplane tyres. *International Journal of Quality & Reliability Management*, 13(8):28–38.

van Hoorn, M. H. and Seelen, L. P. (1983). The $SPP/G/1$ queue: Single server queue with a switched Poisson process as input process. *OR Spektrum*, 5:205–218.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019a). Findings about the two-state BMMPP for modeling point processes in reliability and queueing systems. *Applied Stochastic Models in Business and Industry*, 35(2):177–190.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019b). Fitting procedure for the two-state Batch Markov modulated Poisson process. *European Journal of Operational Research*, 279(1):79–92.

Yuan, F. (2018). Parameter estimation for bivariate weibull distribution using generalized moment method for reliability evaluation. *Quality and Reliability Engineering International*, 34(4):631–640.

Zadeh, A. H. and Bilodeau, M. (2013). Fitting bivariate losses with phase-type distributions. *Scandinavian Actuarial Journal*, 2013(4):241–262.

Zhang, Y., He, X., Liu, Q., Wen, B., and Zheng, J. (2005). Reliability sensitivity of automobile components with arbitrary distribution parameters. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 219(2):165–182.

Zhang, Y. and Liu, Q. (2002). Reliability-based design of automobile components. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 216(6):455–471.

CHAPTER 5

Conclusions and future work

The initial objective of this thesis, as was shown in the Introduction, is to incorporate the *MMPP* point processes into the field of data modeling for which independence in occurrences or exponential distribution between arrival times could not be assumed. The versatility of *MMPP* processes and their extensions allowing group arrivals or in its two-dimensional version, has been highlighted throughout the different chapters included in the dissertation. By way of summary, the work carried out for the estimation of these processes is especially relevant in the thesis. The use of alternative techniques and methodologies has allowed to consider these processes to real problems. As shown in the case studies it has led to some of the methodological advances that have been incorporated into this manuscript.

In this last chapter we compile the main contributions of the thesis and we outline some of the problems in which we think it would be interesting to deepen or advance by taking a further step.

- The Batch Markov modulated Poisson process is studied in Chapter 2 in its general version ($BMMPP_m(K)$), where K is the maximum batch size and m is the number of states of the underlying Markov chain.

  Specifically, the findings of this chapter are listed below and were published in *Yoel G. Yera, Rosa E. Lillo and Pepa Ramírez-Cobo (2019)* Findings about the BMMPP for modeling dependent and simultaneous data in reliability and queueing systems. *Applied Stochastic Models in Business and Industry, 35(2):177-190.*

The most important contributions of this chapter are:

- ✓ The important property of identifiability that holds the $MMPP_m$ is extended to its counterpart with arrivals in batch ($BMMPP_m(K)$).
- ✓ The correlation structure of the two-states Batch Markov modulated Poisson process ($BMMPP_2(K)$) is analyzed.
- ✓ It was illustrated with numerical examples that the correlation between the batch size and the times between events of the $BMMPP_2(K)$ can take both negative and positive values.
- ✓ The non-negativity of the batch size autocorrelation function for the $BMMPP_2(K)$ is proven.
- ✓ These properties makes the $BMMPP_2(K)$ suitable for practical applications.

- Motivated by the results of Chapter 2 in which the identifiability property facilitates the estimation of the $BMMPP_m(K)$, Chapter 3 is devoted to develop the estimation of the process and the validation of the proposed method in a real case.

  This chapter is published in *Yoel G. Yera, Rosa E. Lillo and Pepa Ramírez-Cobo (2019).* Fitting procedure for the two-state Batch Markov modulated Poisson process. *European Journal of Operational Research, 279(1):79-92.*

  The most important contributions of Chapter 3 are the following,

  - ✓ The $BMMPP_2(K)$, represented by $2(K+1)$ parameters, is completely characterized by $2(K+1)$ moments.
  - ✓ An inference approach for fitting real data sets based on a moments matching method is proposed.
  - ✓ The method proposed solves in an iterative way an optimization problem with four unknowns parameters and other $K-1$ optimization problems with two variables.
  - ✓ Both simulated and a real teletraffic trace are used to illustrate the performance of the inference technique proposed.

- ✓ The method proposed is compared to the classic EM algorithm.
- ✓ The proposed algorithm is faster and less dependent on starting points than the EM algorithm.

- A bivariate extension of the two-state Markov modulated Poisson process is introduced in Chapter 4. This novel process inherits the properties of the *MMPP* marginally and also allows for the modeling of correlations between two magnitudes in two-dimensional contexts and allows for the modeling of non-exponential bivariate traces presenting inter and intra-dependence. What is most interesting from our point of view in this chapter is that the modeling is motivated by a real set of data related to train failures that jointly takes into account the time and the kilometers traveled between failures. This paper is submitted for possible publication. The most important contributions of this chapter are:

  - ✓ Some properties of the novel model are show, in particular its identifiability, which is inherited from the marginal processes and is crucial for inference.
  - ✓ A matrix representation for the process that allowed to obtain the joint moments used in the estimation process is obtained.
  - ✓ A fitting approach that combines a matching moment approach with an ABC algorithm is presented.
  - ✓ The methodology proposed is illustrated for both simulated and real data set.

  Obviously, a work of many years is not completely closed in this document. During the time dedicated to understanding this type of process, new paths have come up and they have seemed like problems derived from the initial ideas. Below we list some of these lines of future work that we think can be of interest.

Possible research lines as extensions of the presented work are the following:

- In Chapter 2 it was proved that the autocorrelation function of the $BMMPP_2(K)$ is non-negative. A possible extension of this work is to study the correlation structure of the $BMMPP_m(K)$ for $m \geq 3$ or to find out bounds for the correlation between inter-arrivals times and the batch size for $BMMPP_2(K)$. One of our global concerns is the study of this type of model when the number of states is greater than 2. We think that despite the increase in analytical complexity, the model is much more flexible in modeling real data.

- In Chapter 3 an inference method for the $BMMPP_2(K)$ is proposed, but given that higher order of $BMMPP_m(K)$ are expected to show more versatility for modeling purposes it is of interest to develop inference methods for processes with $m \geq 3$. One option is to develop a moment-matching approach similar to the one proposed in chapter 3 for a $BMMPP_m(K)$ with $m \geq 3$, but it would require finding a set of moments characterizing the processes with higher number of states. Therefore we are facing objectives of both theoretical and algorithmic types.

- In Chapter 4 a bivariate extension of the two-state Markov modulated Poisson process is presented. Because this model has never been studied in the literature, the lines of research emerging are several.

  Some theoretical problems to be considered are:

  - Find closed expressions for quantities of interest as joint probabilities in terms of (T) and (K) or the joint predictive distributions.
  - Obtain probabilities of the counting process as a function of t and k.
  - Extend the process proposed to its batch counterpart, allowing the modeling of situations where failures occur in a simultaneous way.
  - Find a canonical matrix representation that allow an in-depth study of some theoretical properties of this process.

- Obtain a set of moments that characterize the bivariate $MMPP_2$ in order to improve the estimation method of the process.

On the other hand from an applied viewpoint, it is of interest to estimate how many failures will occur in future time and distances intervals, both in terms of expected number and probability of no failure in a given $[t, k]$ interval.

References

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

Akar, N. and Sohraby, K. (2009). System-theoretical algorithmic solution to waiting times in semi-markov queues. *Performance Evaluation*, 66(11):587–606.

Al-Garni, A. Z., Sahin, A. Z., Al-Ghamdi, A. S., and Al-Kaabi, S. A. (1999). Reliability analysis of aeroplane brakes. *Quality and reliability engineering international*, 15(2):143–150.

Ali, S. and Pievatolo, A. (2016). High quality process monitoring using a class of inter-arrival time distributions of the renewal process. *Computers & Industrial Engineering*, 94:45 – 62.

Andersen, A. T. and Nielsen, B. F. (2002). On the use of second-order descriptors to predict queueing behavior of maps. *Naval Research Logistics (NRL)*, 49(4):391–409.

Arts, J. (2017). A multi-item approach to repairable stocking and expediting in a fluctuating demand environment. *European Journal of Operational Research*, 256(1):102–115.

Asmussen, S. (2008). *Applied probability and queues*, volume 51. Springer Science & Business Media.

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

Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian journal of statistics*, 23:419–441.

Assaf, D., Langberg, N. A., Savits, T. H., and Shaked, M. (1984). Multivariate phase-type distributions. *Operations Research*, 32(3):688–702.

Badescu, A. L., Cheung, E. C., and Landriault, D. (2009). Dependent risk models with bivariate phase-type distributions. *Journal of Applied Probability*, 46(1):113–131.

Banerjee, A., Gupta, U., 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.

Banik, A. and Chaudhry, M. (2016). Efficient computational analysis of stationary probabilities for the queueing system $BMAP/G/1/N$ with or without vacation (s). *INFORMS Journal on Computing*, 29(1):140–151.

Bean, N. and Green, D. (1999). When is a MAP poisson? *Mathematical and Computer Modelling*, 82:127–142.

Bearfield, G. and Marsh, W. (2005). Generalising event trees using Bayesian networks with a case study of train derailment. In *International Conference on Computer Safety, Reliability, and Security*, pages 52–66. Springer.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Beirlant, J., Teugels, J. L., and Vynckier, P. (1996). *Practical analysis of extreme values*, volume 50. Leuven University Press Leuven.

Biau, G., Cérou, F., and Guyader, A. (2015). New insights into approximate bayesian computation. In *Annales de l'IHP Probabilités et statistiques*, volume 51, pages 376–403.

Bladt, M. and Nielsen, B. F. (2017). *Matrix-exponential distributions in applied Probability*, volume 81. Springer.

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.

Breuer, L. (2002). An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, 112:123–138.

Buchholz, P. (2003). An EM-algorithm for MAP fitting from real traffic data. In *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 218–236. Springer.

Buchholz, P. and Kriege, J. (2017). Fitting correlated arrival and service times and related queueing performance. *Queueing Systems*, 85(3-4):337–359.

Cai, J. and Li, H. (2005). Multivariate risk model of phase type. *Insurance: Mathematics and Economics*, 36(2):137–152.

Carrizosa, E. and Ramírez-Cobo, P. (2014). Maximum likelihood estimation in the two-state Markovian arrival process. *arXiv preprint arXiv:1401.3105*, Working paper.

Casale, G., Sansottera, A., and Cremonesi, P. (2016). Compact markov modulated models for multiclass trace fitting. *European Journal of Operational Research*, 255(3):822–833.

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

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

Chakravarthy, S. R. (2010). Markovian arrival processes. *Wiley Encyclopedia of Operations Research and Management Science*.

Cordeiro, J. D. and Kharoufeh, J. P. (2011). Batch markovian arrival processes (bmap). *Wiley Encyclopedia of Operations Research and Management Science*.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution*, 25(7):410–418.

Dean, T. A., Singh, S. S., Jasra, A., and Peters, G. W. (2014). Parameter estimation for hidden markov models with intractable likelihoods. *Scandinavian Journal of Statistics*, 41(4):970–987.

Decò, A., Frangopol, D. M., and Zhu, B. (2012). Reliability and redundancy assessment of ships under different operational conditions. *Engineering Structures*, 42:457–471.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Dudin, A. (1998). Optimal multithreshold control for a $BMAP/G/1$ queue with n service modes. *Queueing Systems*, 30(3-4):273–287.

Eliashberg, J., Singpurwalla, N. D., and Wilson, S. P. (1997). Calculating the reserve for a time and usage indexed warranty. *Management Science*, 43(7):966–975.

Ephraim, Y. and Merhav, N. (2002). Hidden markov processes. *IEEE Transactions on information theory*, 48(6):1518–1569.

Ephraim, Y. and Roberts, W. J. (2008). An em algorithm for markov modulated markov processes. *IEEE Transactions on Signal Processing*, 57(2):463–470.

Eum, S., Harris, R., and Atov, I. (2007). A matching model for MAP-2 using moments of the counting process. In *Proceedings of the International Network Optimization Conference, INOC 2007*, Spa, Belgium.

Fang, C. and Das, P. K. (2005). Survivability and reliability of damaged ships after collision and grounding. *Ocean Engineering*, 32(3-4):293–307.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: Semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Fearnhead, P. and Sherlock, C. (2006). An exact Gibbs sampler for the Markov modulated poisson process. *Journal of the Royal Statististical Society: Series B*, 65(5):767–784.

Fischer, W. and Meier-Hellstern, K. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Performance evaluation*, 18(2):149–171.

Ghosh, S. and Banik, A. (2017). An algorithmic analysis of the $BMAP/MSP/1$ generalized processor-sharing queue. *Computers & Operations Research*, 79:1–11.

Goel, L. and Gupta, P. (1984). Analysis of a two-engine aeroplane model with two types of failure and preventive maintenance. *Microelectronics Reliability*, 24(4):663–666.

Green, D. (1998). *$MAP/PH/1$ departure processes*. PhD thesis, School of Applied Mathematics, University of Adelaide, South Australia.

Griffiths, R., Milne, R., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21(3):238–255.

Griffiths, R. C. and Milne, R. K. (1978). A class of bivariate poisson processes. *Journal of Multivariate Analysis*, 8(3):380–395.

He, Q.-M. (2014). *Fundamentals of matrix-analytic methods*, volume 365. Springer.

He, Q.-M. and Zhang, H. (2006). PH-invariant polytopes and coxian representations of phase type distributions. *Stochastic Models*, 22(3):383–409.

He, Q.-M. and Zhang, H. (2008). An algorithm for computing minimal coxian representations. *INFORMS Journal on Computing*, 20:179–190.

He, Q.-M. and Zhang, H. (2009). Coxian representations of generalized Erlang distributions. *Acta Mathematicae Applicatae Sinica. English Series*, 25:489–502.

Heffes, H. and Lucantoni, D. (1986). A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, 4:856–868.

Heindl, A., Mitchell, K., and van de Liefvoort, A. (2006). Correlation bounds for second-order MAPs with application to queueing network decomposition. *Performance Evaluation*, 63(6):553–577.

Heyman, D. and Lucantoni, D. (2003). Modeling multiple IP traffic streams with rate limits. *IEEE/ACM Transactions on Networking*, 11(6):948–958.

Horn, R. A. and Johnson, C. R. (1990). *Matrix analysis*. Cambridge university press.

Horváth, G. and Okamura, H. (2013). A fast EM algorithm for fitting marked Markovian arrival processes with a new special structure. In *European Workshop on Performance Engineering*, pages 119–133. Springer.

Horváth, M., Buchholz, P., and Telek, M. (2005). A map fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. *IEEE/ Second International Conference on the Quantitative Evaluation of Systems (QEST'05)*, pages 124–133.

Horváth, M. and Telek, M. (2002). Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In *Performance evaluation of complex systems: Techniques and Tools, IFIP Performance 2002, in: LNCS Tutorial Series, vol. 2459*, pages 405–434.

Ito, H., Amari, S.-I., and Kobayashi, K. (1992). Identifiability of hidden markov information sources and their minimum degrees of freedom. *IEEE transactions on information theory*, 38(2):324–333.

Ivanov, L. D. (2009). Challenges and possible solutions of the time-variant reliability of ship's hull girder. *Ships and Offshore Structures*, 4(3):215–228.

Jafarian, E. and Rezvani, M. (2012). Application of fuzzy fault tree analysis for evaluation of railway safety risks: an evaluation of root causes for passenger train derailment. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226(1):14–25.

Jatiningsih, L., Respatiwulan, Susanti, Y., Handayani, S. S., and Hartatik (2019). The parameter estimation of conditional intensity function temporal point process as renewal process using bayesian method and its application on the data of earthquake in east nusa tenggara. *Journal of Physics: Conference Series*, 1217:12–63.

Kang, S. H., Kim, Y. H., Sung, D. K., and Choi, B. D. (2002). An application of markovian arrival process (MAP) to modeling superposed atm cell streams. *IEEE Transactions on Communications*, 50(4):633–642.

Kang, S. H. and Sung, D. K. (1995). Two-state $MMPP$ modeling of ATM superposed traffic streams based on the characterization of correlated interarrival times. In *Proceedings of GLOBECOM'95*, volume 2, pages 1422–1426. IEEE.

Karim, M. R. (2008). Modelling sales lag and reliability of an automobile component from warranty database. *International Journal of Reliability and Safety*, 2(3):234–247.

Kim, C. S., Klimenok, V., Mushko, V., and Dudin, A. (2010). The bmap/ph/n retrial queueing system operating in markovian random environment. *Computers & Operations Research*, 37(7):1228–1237.

Kızılaslan, F. and Nadar, M. (2018). Estimation of reliability in a multicomponent stress–strength model based on a bivariate Kumaraswamy distribution. *Statistical Papers*, 59(1):307–340.

Klemm, A., Lindemann, C., and Lohmann, M. (2003). Modeling IP traffic using Batch Markovian Arrival Process. *Performance Evaluation*, 54(2):149–173.

Kriege, J. and Buchholz, P. (2011). Correlated phase-type distributed random numbers as input models for simulations. *Performance Evaluation*, 68(11):1247–1260.

Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operations research*, 37(1):151–158.

Kypraios, T., Neal, P., and Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical biosciences*, 287:42–53.

Landon, J., Özekici, S., and Soyer, R. (2013). A markov modulated poisson model for software reliability. *European Journal of Operational Research*, 229(2):404–410.

Landriault, D. and Shi, T. (2015). Occupation times in the map risk model. *Insurance: Mathematics and Economics*, 60:75–82.

Latouche, G. and Ramaswami, V. (1999). *Introduction to matrix analytic methods in stochastic modeling*, volume 5. SIAM.

Li, M., Chen, W., and Han, L. (2010). Correlation matching method for the weak stationarity test of lrd traffic. *Telecommunication Systems*, 43(3-4):181–195.

Li, S. and Ren, J. (2013). The maximum severity of ruin in a perturbed risk process with markovian arrivals. *Statistics & Probability Letters*, 83(4):993–998.

Liu, B., Cui, L., Wen, Y., and Shen, J. (2015). A cold standby repairable system with working vacations and vacation interruption following markovian arrival process. *Reliability Engineering & System Safety*, 142:1–8.

Lucantoni, D. (1991). New results for the single server queue with a Batch Markovian Arrival Process. *Stochastic Models*, 7:1–46.

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

Lucantoni, D., Choudhury, G., and Whitt, W. (1994). The transient $BMAP/G/1$ queue. *Stochastic Models*, 10:145–182.

Lucantoni, D., Meier-Hellstern, K., and Neuts, M. (1990). A single-server queue with server vacations and a class of nonrenewal arrival processes. *Advances in Applied Probability*, 22:676–705.

Lveill, G. and Hamel, E. (2019). Compound trend renewal process with discounted claims: a unified approach. *Scandinavian Actuarial Journal*, 2019(3):228–246.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian Computational methods. *Statistics and Computing*, 22(6):1167–1180.

Marshall, A. H., Mitchell, H., and Zenga, M. (2015). *Modelling the Length of Stay of Geriatric Patients in Emilia Romagna Hospitals Using Coxian Phase-Type Distributions with Covariates*, pages 127–139. Springer International Publishing, Cham.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, 110(4):1321–1326.

Minter, A. and Retkute, R. (2019). Approximate Bayesian Computation for infectious disease modelling. *Epidemics*, page 100368.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2006). Reliability of a system under two types of failures using a Markovian arrival process. *Operations Research Letters*, 34:525–5530.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2014a). Matrix stochastic analysis of the maintainability of a machine under shocks. *Reliability Engineering & System Safety*, 121:11–17.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2014b). A redundant n-system under shocks and repairs following markovian arrival processes. *Reliability Engineering & System Safety*, 130:69–75.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2015). A reliability system under cumulative shocks governed by a bmap. *Applied Mathematical Modelling*, 39(23-24):7620–7629.

Montoro-Cazorla, D., Pérez-Ocón, R., and Segovia, M. (2009). Replacement policy in a system under shocks following a Markovian arrival process. *Reliability Engineering and System Safety*, 94:497–502.

Nadar, M. and Kızılaslan, F. (2015). Estimation of reliability in a multicomponent stress-strength model based on a Marshall-Olkin bivariate Weibull distribution. *IEEE Transactions on Reliability*, 65(1):370–380.

Nakagawa, T. (2011). *Stochastic processes: With applications to reliability theory*. Springer Science & Business Media.

Narayana, S. and Neuts, M. (1992). The first two moment matrices of the counts for the Markovian arrival process. *Communications in statistics. Stochastic models*, 8(3):459–477.

Nasr, W. W., Charanek, A., and Maddah, B. (2018). Map fitting by count and inter-arrival moment matching. *Stochastic Models*, pages 1–29.

Navas, M. A., Sancho, C., and Carpio, J. (2017). Reliability analysis in railway repairable systems. *International Journal of Quality & Reliability Management*, 34(8):1373–1398.

Neuts, M. and Li, J. (1997). *An algorithm for the $P(n, t)$ matrices of a continuous BMAP*, volume 183 of *Lectures notes in Pure and Applied Mathematics*, pages 7–19. Srinivas R. Chakravarthy and Attahiru, S. Alfa, editors. NY: Marcel Dekker, Inc.

Neuts, M. F. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779.

Okamura, H. and Dohi, T. (2009). Faster maximum likelihood estimation algorithms for markovian arrival processes. In *2009 Sixth international conference on the quantitative evaluation of systems*, pages 73–82. IEEE.

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

Okamura, H., Dohi, T., and Trivedi, K. (2011). A refined em algorithm for ph distributions. *Performance Evaluation*, 68(10):938–954.

Okamura, H., Kishikawa, H., and Dohi, T. (2013). Application of deterministic annealing em algorithm to map/ph parameter estimation. *Telecommunication Systems*, 54(1):79–90.

Özekici, S. and Soyer, R. (2003). Reliability of software with an operational profile. *European Journal of Operational Research*, 149(2):459–474.

Özekici, S. and Soyer, R. (2006). Semi-markov modulated poisson process: probabilistic and statistical analysis. *Mathematical Methods of Operations Research*, 64(1):125–144.

Pievatolo, A. and Ruggeri, F. (2010). Bayesian modelling of train door reliability. *The Oxford Handbook of Applied Bayesian Analysis. Oxford University Press, Oxford*, pages 271–294.

Pievatolo, A., Ruggeri, F., and Argiento, R. (2003). Bayesian analysis and prediction of failures in underground trains. *Quality and Reliability Engineering International*, 19(4):327–336.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798.

Ramaswami, V. (1990). From the matrix-geometric to the matrix-exponential. *Queueing Systems*, 6:229–260.

Ramírez-Cobo, P. and Carrizosa, E. (2012). A note on the dependence structure of the two-state Markovian arrival process. *Journal of Applied Probability*, 49:295–302.

Ramírez-Cobo, P. and Lillo, R. (2012). New results about weakly equivalent MAP_2 and MAP_3 processes. *Methodology and Computing in Applied Probability*, 14(3):421–444.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2008). Bayesian analysis of a queueing system with a long-tailed arrival process. *Communications in StatisticsSimulation and Computation*, 37(4):697–712.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2010). Nonidentifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47(3):630–649.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2014a). Identifiability of the $MAP_2/G/1$ queueing system. *Top*, 22(1):274–289.

Ramírez-Cobo, P., Lillo, R., and Wiper, M. (2017). Bayesian analysis of the stationary MAP_2. *Bayesian Analysis*, 12(4):1163–1194.

Ramírez-Cobo, P., Lillo, R. E., Wilson, S., and Wiper, M. P. (2010). Bayesian inference for Double Pareto lognormal queues. *Annals of Applied Statistics*, 4(3):1533–1557.

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

Revzina, E. (2010). Stochastic models of data flows in the telecommunication networks. *Computer Modelling and New Technologies*, 14(2):29–34.

Riska, A., Diev, V., and Smirni, E. (2004). Efficient fitting of long-tailed data sets into phase-type distributions. *Performance Evaluation Journal*, 55:147–164.

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 and System Safety*, 134:126–133.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016a). Analytical issues regarding the lack of identifiability of the non-stationary MAP_2. *Performance Evaluation*, 102:1–20.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016b). Dependence patterns for modeling simultaneous events. *Reliability Engineering and System Safety*, 154:19–30.

Rodríguez, J., Lillo, R., and Ramírez-Cobo, P. (2016c). Nonidentifiability of the two-state $BMAP$. *Methodology and Computing in Applied Probability*, 18(1):81–106.

Ruggeri, F. (2006). On the reliability of repairable systems: methods and applications. In *Progress in Industrial Mathematics at ECMI 2004*, pages 535–553. Springer.

Rydén, T. (1994). Parameter estimation for Markov modulated Poisson processes. *Stochastic Models*, 10(4):795–829.

Rydén, T. (1996a). An EM algorithm for estimation in Markov modulated Poisson processes. *Computational Statistics & Data Analysis*, 21(4):431–447.

Rydén, T. (1996b). An EM algorithm for estimation in Markov modulated Poisson processes. *Computational Statistics & Data Analysis*, 21(4):431–447.

Rydén, T. (1996c). 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.

Scott, S. (1999). Bayesian analysis of the two state markov modulated poisson process. *Journal of Computational and Graphical Statistics*, 8(3):662–670.

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.

Sheikh, A. K., Al-Garni, A. Z., and Affan Badar, M. (1996). Reliability analysis of aeroplane tyres. *International Journal of Quality & Reliability Management*, 13(8):28–38.

Sikdar, K. and Samanta, S. (2016). Analysis of a finite buffer variable batch service queue with batch markovian arrival process and servers vacation. *Opsearch*, 53(3):553–583.

Singh, G., Gupta, U., and Chaudhry, M. (2016). Detailed computational analysis of queueing-time distributions of the $BMAP/G/1$ queue using roots. *Journal of Applied Probability*, 53(4):1078–1097.

Takine, T. (2016). Analysis and computation of the stationary distribution in a special class of markov chains of level-dependent $M/G/1$-type and its application to $BMAP/M/\infty$ and $BMAP/M/c+M$ queues. *Queueing Systems*, 84(1-2):49–77.

Telek, M. and Horváth, G. (2007). A minimal representation of markov arrival processes and a moments matching method. *Performance Evaluation*, 64(9-12):1153–1168.

van Hoorn, M. H. and Seelen, L. P. (1983). The $SPP/G/1$ queue: Single server queue with a switched Poisson process as input process. *OR Spektrum*, 5:205–218.

Wang, K., Tao, M., Chen, W., and Guan, Q. (2015). Delay-aware energy-efficient communications over nakagami-m fading channel with $MMPP$ traffic. *IEEE Transactions on Communications*, 63(8):3008–3020.

Wilkinson, R. D. (2013). Approximate bayesian computation (abc) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, 12(2):129–141.

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*, 35:1184–1193.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019a). Findings about the two-state BMMPP for modeling point processes in reliability and queueing systems. *Applied Stochastic Models in Business and Industry*, 35(2):177–190.

Yera, Y. G., Lillo, R. E., and Ramírez-Cobo, P. (2019b). Fitting procedure for the two-state Batch Markov modulated Poisson process. *European Journal of Operational Research*, 279(1):79–92.

Yuan, F. (2018). Parameter estimation for bivariate weibull distribution using generalized moment method for reliability evaluation. *Quality and Reliability Engineering International*, 34(4):631–640.

Zadeh, A. H. and Bilodeau, M. (2013). Fitting bivariate losses with phase-type distributions. *Scandinavian Actuarial Journal*, 2013(4):241–262.

Zhang, Y., He, X., Liu, Q., Wen, B., and Zheng, J. (2005). Reliability sensitivity of automobile components with arbitrary distribution parameters. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 219(2):165–182.

Zhang, Y. and Liu, Q. (2002). Reliability-based design of automobile components. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 216(6):455–471.