Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

(Batch) Markovian arrival processes: the identifiability issue and other applied aspects

Autor:

Joanna Virginia Rodríguez César

Directores:

Rosa E. Lillo

Pepa Ramírez-Cobo

DEPARTAMENTO DE ESTADÍSTICA

Leganés, Marzo de 2015

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

(Batch) Markovian arrival processes: the identifiability issue and other applied aspects

Autor:

Joanna Virginia Rodríguez César

Director/es:

Rosa E. Lillo

Pepa Ramírez-Cobo

Firma del Tribunal Calificador:

Firma

Presidente: Rafael Pérez Ocón.

Vocal: Mogens Bladt

Secretario: Bernardo D'Auria

Calificación:

Leganés, 26 de Marzo de 2015

TESIS DOCTORAL

(Batch) Markovian arrival processes: the identifiability issue and other applied aspects

Joanna Virginia Rodríguez César

Advisers: Rosa E. Lillo and Pepa Ramírez-Cobo

Department of Statistics

UNIVERSIDAD CARLOS III DE MADRID

Leganés, March 2015

To my family

Contents

Acknowledgements

First of all, I want to express all my gratitude to all those people that have been a part of this journey in one way or another.

I would specially like to thank my advisers Dr. Rosa Elvira Lillo and Dr. Pepa Ramírez Cobo for their help, support and guidance over the past four years. We have been a team, and this achievement has been possible thanks to their dedication, patience, encouragement and enthusiasm. Not only have I learned about the research topic of this dissertation, but also their personal experiences helped me a great deal when having to make important decisions. In a nutshell, it would not be possible without them.

I also want to express my gratitude to Dr. Tobias Rydén with whom I worked closely during the spring of 2012 in Stockholm. His knowledge and experience have helped me to better understand many theoretical and computational problems. I am very thankful for his time, patience, criticism and friendship.

To my highschool math teacher, Alcides, thanks for introducing me to the wonderful world of Mathematics. To my university professors, Dr. Claudia Antonini and Dr. Alfredo Ríos, who were strong pillars in my training. Thanks to them I discovered probability and statistics, as well as great persons who have guided and support me.

I am very thankful to all members and staff of the Statistics Department at University Carlos III of Madrid, for always being kind, supportive, helpful and friendly towards me. I am extremely grateful for the people I have met and the friendships I have made in the Department. You have all been a part of this journey, and have helped to make it less difficult.

And finally, but surely no less important, I want to thank my family and friends back home, for always supporting me in this life decision that separated me from my home and country. Venezuela is always in my heart. Hope you are all proud.

Abstract

This dissertation is mainly motivated by the problem of statistical modeling via a specific point process, namely, the Batch Markovian arrival processes. Point processes arise in a wide range of situations of our daily activities, such as people arriving to a bank, claims of an insurance company or failures in a system. They are defined by the occurrence of an event at a specific time, where the event occurrences may be understood from different perspectives, either by the arrival of a person or group of people in a waiting line, the different claims to the insurance companies or failures occurring in a system. Point processes are defined in terms of one or several stochastic processes which implies more versatility than mere single random variables, for modeling purposes.

A traditional assumption when dealing with the analysis of point processes is that the occurrence of events are independent and identically distributed, which considerably simplifies the theoretical calculations and computational complexity, and again because of simplicity, the Poisson process has been widely considered in stochastic modelling. However, the independence and exponentiability assumptions become unrealistic and restrictive in practice. For example, in teletraffic or insurance contexts it is usual to encounter dependence amongst observations, high variability, arrivals occurring in batches, and therefore, there is a need of more realistic models to fit the data.

In particular, in this dissertation we investigate new theoretical and applied properties concerning the (batch) Markovian arrival processes, or $(B)MAP$, which is well known to be a versatile class of point process that allows for dependent and non-exponentially distributed inter-event times as well as correlated batches. They inherit the tractability of the Poisson processes, and turn out suitable models to fit data with statistical features that differ form the classical Poisson assumptions. In addition, in spite of the large amount of works considering the $BMAP$, still there are a number of open problems which are of interest and which shall be considered in this dissertation.

This dissertation is organized as follows. In Chapter 1, we present a brief theoretical background that introduces the most important concepts and properties that are needed to carry out our analyses. We give a theoretical background of point processes and describe them from a probabilistic point of view. We introduce the Markovian point processes and its main properties, and also provide some point process estimation backdrop with a review of recent works.

An important problem to consider when the statistical inference for any model is to be developed is the uniqueness of its representation, the identifiability problem. In Chapter 2 we analyze the identifiability of the non-stationary two-state *MAP*. We prove that, when the sample information is given by the inter-event times, then, the usual parametrization of the process is redundant, that is, the process is nonidentifiable. We present a methodology to build an equivalent non-stationary two-state *MAP*s from any fixed one. Also, we provide a canonical and unique parametrization of the process so that the redundant versions of the same process can be reduced to its canonical version.

In Chapter 3 we study an estimation approach for the parameters of the non-stationary version of the *MAP* under a specific observed information. The framework to be considered is the modelling of the failures of N electrical components that are identically distributed, but for which it is not reasonable to assume that the operational times related to each component are independent and identically distributed. We propose a moments matching estimation approach to fit the data to the non-stationary two-state *MAP*. A simulated and a real data set provided by the Spanish electrical group Iberdrola are used to illustrate the approach.

Unlike Chapters 2 and 3, which are devoted to the Markovian arrival process, Chapters 4 and 5 focus on its arrivals-in-batches counterpart, the *BMAP*. The capability of modeling non-exponentially distributed and dependent inter-event times as well as correlated batches makes the *BMAP* suitable in different real-life settings as teletraffic, queueing theory or actuarial contexts, to name a few. In Chapter 4 we analyze the identifiability issue of the *BMAP*. Specifically, we explore the identifiability of the stationary two-state *BMAP* noted as $BMAP_2(k)$, where k is the maximum batch arrival size, under the assumptions that both the inter-event times and batches sizes are observed. It is proven that for $k \geq 2$ the process cannot be identified. The proof is based on the construction of an equivalent $BMAP_2(k)$ to a given one, and on the decomposition of a $BMAP_2(k)$ into k $BMAP_2(2)$s.

In Chapter 5 we study the auto-correlation functions of the inter-event times and batch sizes of the *BMAP*. This chapter examines the characterization of both auto-correlation functions for the stationary $BMAP_2(k)$, for $k \geq 2$, where four behavior patterns are identified

for both functions for the $BMAP_2(2)$. It is proven that both auto-correlation functions decrease geometrically as the time lag increases. Also, the characterization of the auto-correlation functions has been extended for the general $BMAP_m(k)$ case, $m \geq 3$.

To conclude, Chapter 6 summarizes the most significant contributions of this dissertation, and also give a short description of possible research lines.

Resumen

Esta tesis está motivada por el problema de modelización estadística mediante un tipo específico de procesos puntuales, los procesosde llegada Markovianos en tandas. Los procesos puntuales surgen en una gran variedad de situaciones de la vida real, como las personas que llegan a un banco, reclamaciones en compañías de seguro o fallos en un sistema. Los procesos puntuales se definen como la ocurrencia de eventos en diferentes instantes temporales, donde las ocurrencias de eventos se pueden entender desde diferentes perspectivas, llegadas de personas o un grupo de personas a una cola, las distintas reclamaciones en una compañía de seguros o los fallos que ocurren en un sistema. Los procesos puntuales se definen en términos de uno o varios procesos estocásticos lo que implica más versatilidad, en términos de modelización, que la que se obtiene mediante variables aleatorias que no consideren la dimensión temporal.

Una suposición tradicional en la literatura al estudiar y analizar procesos puntuales es que los tiempos entre la ocurrencia de eventos son independientes e idénticamente distribuidos, lo que simplifica considerablemente los cálculos teóricos y la complejidad computacional. Adicionalmente, por simplicidad, el proceso de Poisson ha sido ampliamente considerado en modelización estocástica. Sin embargo, las suposiciones de independencia y exponenciabilidad son poco realistas en la práctica. Por ejemplo, en el contexto teletráfico o de seguros es usual encontrar dependencia entre las observaciones, alta variabilidad, llegadas que ocurren en tandas, por lo que hay una necesidad de ajustar los datos a modelos más reales.

En particular, en esta tesis investigamos nuevas propiedades teóricas y aplicadas sobre los procesos de llegada Markovianos (en tanda), denotados $(B)MAP$, que son conocidos por ser procesos puntuales versátiles que permiten la dependencia y no-exponenciabilidad de los tiempos entre eventos, así como la correlación entre las tandas. Ya que heredan la manejabilidad de los procesos de Poisson, son procesos adecuados para ajustar datos con características estadísticas que difieren de los supuestos clásicos de Poisson. Además, a pesar de la gran cantidad de trabajos que consideran los $BMAP$, todavía hay una serie de

problemas abiertos que son de interés y que serán considerados en esta tesis.

La estructura de esta tesis es la siguiente. En el Capítulo 1, se presenta una breve revisión teórica que introduce las definiciones y propiedades más importantes necesarias para el desarrollo de nuestros análisis. Se definen los procesos puntuales y se describen desde un punto de vista probabilístico. Se introducen los procesos puntuales Markovianos y sus propiedades principales, además se proporciona una revisión de la literatura sobre la estimación de los procesos puntuales.

Un problema importante a considerar cuando se quieren desarrollar métodos de inferencia sobre cualquier modelo es la unicidad de su parametrización, o alternativamente, el problema de identificabilidad. En el Capítulo 2 estudiamos el problema de identificabilidad del MAP no estacionario con dos estados. Se demuestra que, cuando la información muestral está dada por los tiempos entre eventos, entonces, la parametrización usual del proceso es redundante, esto es, el proceso es no-identificable. Se presenta un procedimiento para construir un MAP no estacionario con dos estados equivalente a uno fijo. Además, se proporciona una parametrización canónica y única del proceso, de manera que las versiones redundantes o equivalentes de un mismo proceso se pueden reducir a su versión cnónica.

En el Capítulo 3 se estudia un método de estimación para los parámetros del MAP no estacionario con dos estados. El esquema que se considerará es la modelización de los fallos de N componentes eléctricos que son idénticamente distribuidos, pero que no es razonable considerar que los tiempos operacionales asociados a cada componente son independientes ni idénticamente distribuidos. Se propone un método de igualdad de momentos para ajustar datos a un MAP no estacionarios con dos estados. Se presenta un ejemplo simulado y un ejemplo con datos reales proporcionados por la compañía eléctrica Iberdrola para ilustrar la metodología propuesta.

A diferencia de los capítulos 2 y 3, que están dedicados a los procesos de llegada Markovianos, los capítulos 4 y 5 se centran en su generalización para considerar llegadas en tandas, el $BMAP$. La capacidad de modelar tiempos entre eventos dependientes y no-exponenciales, así como llegadas en tandas correladas, hace que los $BMAP$ sean modelos apropiados en problemas de la vida real, como en contextos teletráficos, de teoría de colas o actuariales, entre otros. En el Capítulo 4 se explora la identificabilidad para el $BMAP$ estacionario de 2 estados, $BMAP_2(k)$, donde k es el tamaño máximo de las tandas, bajo la suposición de que los tiempos entre eventos y los tamaños de las tandas son los datos observados. Se demuestra que para $k \geq 2$ el proceso no es único. La demostración se basa en la construcción de un $BMAP_2(k)$ equivalente a uno fijo, y en la descomposición de un $BMAP_2(k)$ en k $BMAP_2(2)$s.

En el Capítulo 5 se estudia las funciones de autocorrelación para los tiempos entre-eventos y las llegadas en tanda del *BMAP*. Además, también se examina la caracterización de ambas funciones de autocorrelación para el $BMAP_2(k)$, $k \geq 2$, estacionario, donde se identifican cuatro patrones para el $BMAP_2(2)$. Se demuestra que ambas funciones de autocorrelación decrecen geométricamente. Finalmente, se extiende la caracterización de las funciones de autocorrelación para el caso general $BMAP_m(k)$, $m \geq 3$.

Finalmente, en el Capítulo 6 se resumen las contribuciones más importantes de esta tesis y futuras líneas de investigación.

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces point processes and the main properties concerning the (Batch) Markovian arrival process, the model we focus on in this dissertation. Some real data examples from the insurance and reliability contexts are shown to motivate the considered research. Special emphasis is put on the description of the phase-type distribution as well as the Markovian renewal theory, both needed to later define the *BMAP*. Finally, the chapter provides a review of the classic statistical estimation approaches for inference of point processes, where the identifiability issue is of crucial importance.

1.1 Point processes and real data

Point processes are defined as the occurrence of events at different instant epochs, where the *occurrence of event* is defined depending on the context. For example, in teletraffic an event may denote the arrival of a packet of bytes; in insurance it may indicate the occurrence of any type of risk; in reliability, an event may be understood as a failure system and in queuing theory, the arrival of a customer. Some real data examples of these situations are presented next. From the teletraffic context, the publicly available Bellcore LAN trace files, named BC-pAug89 are found in

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

The data file consists of two columns in ASCII format, where the first column gives the time in seconds of the packet arrival, and the second column gives the Ethernet data length in bytes, shown in Figure 1.1 for the first 100 packet arrivals. The trace began at 11:25 on August 29, 1989, and ran for about 3142.82 seconds (until 1,000,000 packets had been captured) at the Bellcore Morristown Research and Engineering facility. The next example is from an insurance context: a total of 600 claims and their corresponding amounts, provided by the insurance department of an international commercial company (see Vilar et al. [104]), shown in Figure 1.2. Finally, consider an example from a reliability context: the failure times of 926 energy generators (electrical components) provided by the Spanish private electrical utility company, Iberdrola, shown in Figure 1.3.

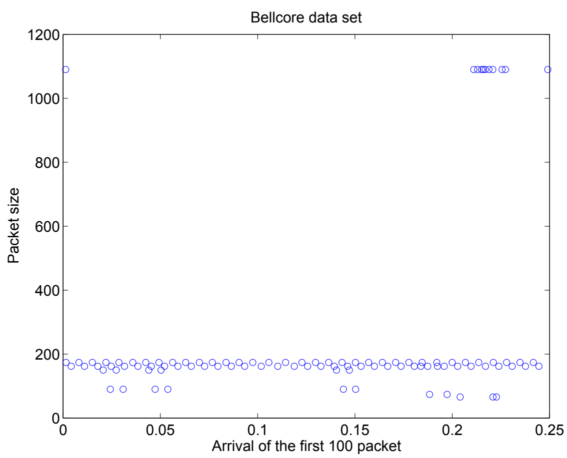

Figure 1.1: Arrival time of 100 packets.

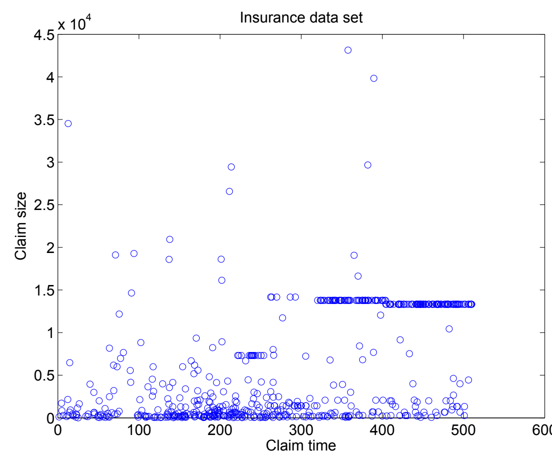

Figure 1.2: Occurrence time of 600 claims.

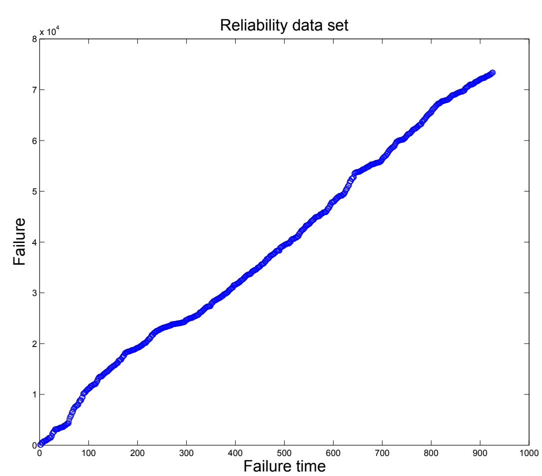

Figure 1.3: Occurrence time of 926 failures of an electrical component.

Consider the problem of statistically modeling real inter-event times. Then, the Poisson process could be thought as a first approach because of its tractability. However, the hypothesis of independent and exponentially distributed inter-event times might not be realistic in practice, as the following set of figures illustrates

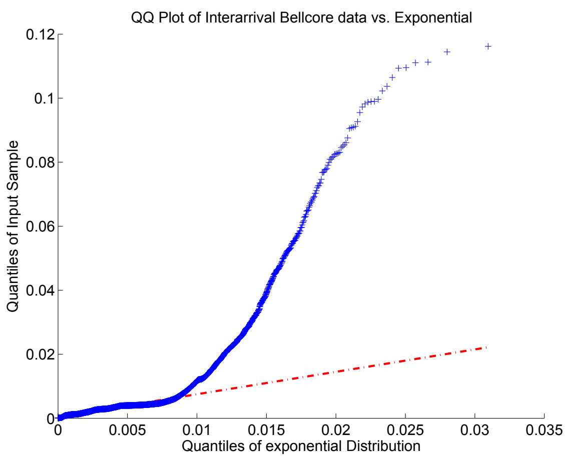

Figure 1.4: Packets interarrival times QQ-plot.

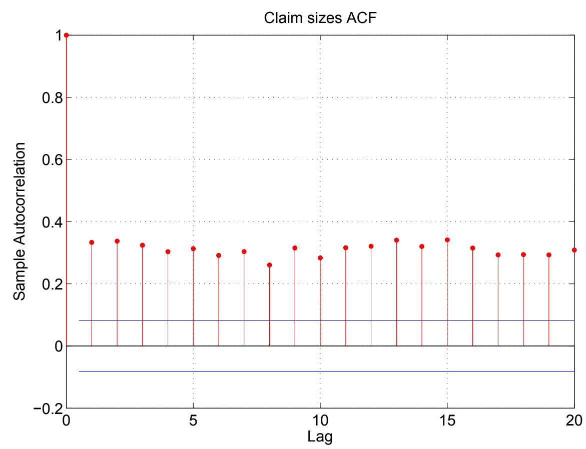

Figure 1.5: Claim sizes empirical ACF.

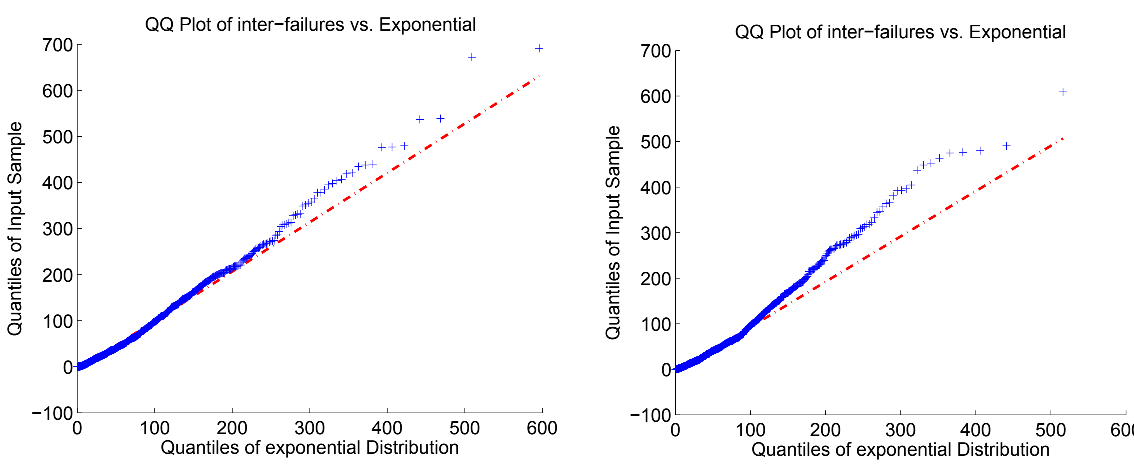

Figure 1.6: Reliability data set. Left panel: QQ-plot of the inter-failures until the first failure. Right panel: QQ-plot of the inter-failures between the first and second failures.

Figure 1.4 shows the QQ-plot of the packet interarrival times from the teletraffic example, in comparison with that of an exponential distribution. It is clear that the exponential distribution would not provide a good fit for the data. Let us analyze Figure 1.5 which depicts the empirical auto-correlation function for the amount of the claims, from the insurance example data set. We can observe that the claim sizes has a correlated structure, in consequence, a model that assumes independence between the data would not be appropriate. Finally, in Figure 1.6, the left and right panels depict the QQ-plots of the inter-failure times until the first failure and between the first and second failures (of all the generators) respectively, compared to an exponential distribution. As before, the exponential distribution does not perform well. As the previous examples in different contexts have shown, there is a need for appropriate point processes that properly captures these statistical features.

Regarding statistical dependence properties and non-exponentiability of the data, the (Batch) Markovian arrival processes, $(B)MAP$ ([62, 75]), play an important part in the stochastic modeling world, from both a theoretical and a practical point of view, since it allows for dependent and non- exponentially distributed inter-event times. In addition, it may also include (correlated) event occurrences in batches. This versatility makes the $(B)MAP$ a suitable point process for modeling real-life situations and indeed, there is a large amount of works dealing with both theoretical and applied aspects of the process. In this dissertation we aim to explore theoretical and applied properties concerning the $(B)MAP$.

1.2 Point processes

In this section a formal definition of point processes is given as well as the description of the elements in the stochastic theory needed to define the *(B)MAP*.

A point process is a random sequence of epochs at which a certain event occurs. Figure 1.7 illustrates how a point process behaves, where $\{S_n, n \geq 1\}$, with $S_0 = 0$, denote an increasing sequence of random variables that represent the time of the n-th event occurrence (or epoch times), where simultaneous occurrences of groups or batches are allowed. The elapsed time between S_{n-1} and S_n is denoted by T_n, i.e., the random variable $T_n = S_n - S_{n-1}$ represents the inter-event times. The variable S_n can be written as a function of T_n, as

$$S_n = \sum_{i=1}^{n} T_i.$$

A point process can also be specified by the counting process $\{N(t), t \geq 0\}$, which is a random variable that represents the number of events that have occurred in the interval

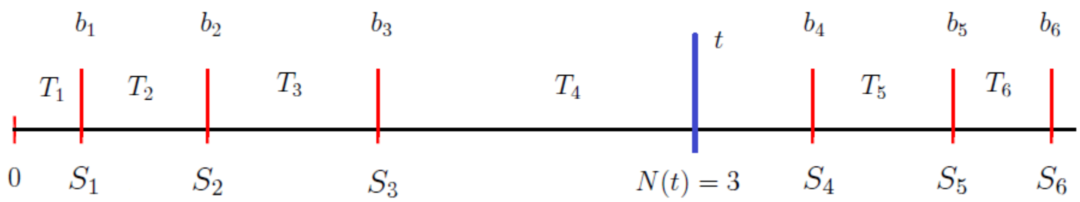

Figure 1.7: Point process where S_n are the event occurrence times and T_n its inter-event times. The values b_i denotes the batch size of the i-th event occurrence.

$(0, t]$, that is,

$$N(t) = \max\{n : S_n \leq t\} \quad t \geq 0.$$

Note that,

$$\{S_n \leq t\} \quad \text{if and only if} \quad \{N(t) \geq n\}, \quad \text{for } n, t \geq 1,$$

that is, $\{S_n \leq t\}$ is the occurrence of the n-th event at most at time t, which implies that the number of events is n at time t, therefore $\{N(t) \geq n\}$. The converse is direct. Then, a point process can be specified by the times of the event times, the inter-event times or by the counting process. Specifying one, generally, the other can be specified as well.

1.2.1 Renewal Process

Renewal process are an important type of point processes for which the sequence of inter-event times are independent and identical distributed random variables, with an arbitrary distribution. That is, the event occurrence process, S_n, is a sum of non-negative i.i.d. random variables (the inter-event times). These processes are called renewal process because the process probabilistically starts over at each event occurrence, independently of the past. One of the main reasons to study renewal processes is that many complicated processes have randomly occurring instants at which the system returns to a state probabilistically equivalent to the starting state. These renewal epochs allow us to separate the long-term behavior of the process that can be studied through renewal theory from the behavior within each renewal period. Let us recall that a counting process is said to possess independent increments if the number of occurrences that occur in disjoint time intervals are independent. It is said to possess stationary increments if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. The most common and well known counting process is the Poisson process defined as,

Definition 1.1. *A counting process* $\{N(t), t \geq 0\}$ *is called a Poisson process if*

1. $N(0) = 0$.
2. $\{N(t), t \geq 0\}$ *has the independent increment property.*
3. $P[N(s+t) - N(s) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$, *for* $n \geq 0$. *That is, the number of event occurrences in any interval of length* t *has a Poisson distribution with mean* λt.

Condition 3 implies that the distribution of $N(s, s+t)$ is independent from t, that is, the process $\{N(t), t \geq 0\}$ has stationary increments. It also indicates that λ is the expected number of events per unit time in the Poisson process. The events in a Poisson process are called arrivals, therefore λ is called the arrival rate of the process. If the parameter λ is independent of time, the Poisson process is called a homogeneous Poisson process. On the contrary, if λ is a function of time, $\lambda(t)$, then such processes are called non-homogeneous Poisson process. In this dissertation we are mainly concerned with homogeneous processes. For a Poisson process, let the random variable T_n be defined as before, the sequence of inter-event times. The following result determines the distribution of T_n (see Chapter 9, Ross [99]).

Proposition 1.1. *The inter-event times* $\{T_n, n \geq 1\}$ *are independent and identically distributed exponential random variables, with mean* $1/\lambda$.

The assumption of independent and identically distributed exponential inter-event times is highly restrictive in practice. For this reason, Neuts [74] developed the theory of phase-type distributions, which are presented as a natural generalization of the exponential distribution.

1.2.1.1 Phase-type distributions

Phase-type renewal processes are renewal processes for which the sequence of inter-event times are independent and identical distributed phase-type random variables. Phase-type (PH) distributions play an important role when defining the process that is going to be developed in this dissertation, the *(B)MAP*. The continuous PH-distributions are a generalization of the exponential distribution and define a more versatile class of distributions. They were introduced by Neuts [74] as an alternative distribution when the Poisson/exponential distributions are not appropriate models.

The main purpose of using Poisson/exponential distributions in modeling is due to their memoryless property, which leads to tractable results. *PH*-distributions generalize the exponential distribution and constitute a flexible class of probability models. They have been widely used in practice in many areas, since they are analytically and algorithmically tractable models. For example, in reliability (see Montoro-Cazorla et al. [68] and Peng et al. [82]), in queueing theory (see Chakravarthy and Neuts [18] and Kim et al. [50]), and in healthcare (see Marshall et al. [66] and Gillespie et al. [33]). Let us recall the exponential distribution definition and its main properties.

Definition 1.2. *Let X be a non-negative random variable. X is an exponential random variable with parameter $\lambda > 0$, if its cumulative distribution function is given as*

$$F_X(t) = P[X \leq t] = 1 - exp(-\lambda t) = 1 - \sum_{n=0}^{\infty} \frac{(-\lambda)^n t^n}{n!}, \qquad t \geq 0.$$

The parameter $\lambda > 0$ is the arrival rate of the Poisson process. The most important property of the exponential distribution is that it is the only continuous distribution with the memoryless property:

$$P[X > t + s | X > t] = P[X > s], \qquad t, s \geq 0,$$

which states that the chance of an event occurring does not depend on the elapsed time. It is because of the memoryless property that Markov processes are used as models.

A *PH*-distribution can be described as a mixture of exponential distributions, each representing a phase (or state), with or without the same rate parameter. Consider a process with m phases, after it starts in phase i (for $i = 1, \ldots, m$), the processes jumps to phase j (for $j = 1, \ldots, m$), and so on until the transitions stop and the process ends, with an exponentially distributed sojourn time in each phase. The *PH* random variable is described as the total time spent jumping through the exponential phases until the process ends. Following He [36], an algebraic definition of the *PH*-distribution is as follows.

Definition 1.3. *Let X be a non-negative random variable. X is is said to be PH-distributed, represented as $(\boldsymbol{\alpha}, T)$, if its cumulative distribution function is given as*

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

where

i) $\mathbf{e}$ *is a column vector of ones.*

ii) $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ *is a row vector of order* $m > 0$ *(phases), where* $\alpha_i \geq 0$ *and* $\boldsymbol{\alpha}\mathbf{e} = 1$.

iii) T *is an* $m \times m$ *matrix that satisfies*

(a) $(T)_{ii} < 0$,

(b) $(T)_{ij} \geq 0$,

(c) *all row sums are non-positive, and*

(d) T *is invertible*

Some basic distributional characteristics of a PH-random variable X with representation $(\boldsymbol{\alpha}, T)$, are

1. The density function is,

$$f_X(s) = \boldsymbol{\alpha} e^{Ts}(-T\mathbf{e}), \quad s \geq 0.$$

2. The Laplace transform is

$$f_X^*(s) = \boldsymbol{\alpha}\left(sI - T\right)^{-1}(-T\mathbf{e}), \quad s \geq 0, \quad \text{where } I \text{ is the identity matrix.}$$

3. The i-th moment is given by,

$$E\left(X^i\right) = i!\boldsymbol{\alpha}\left(-T\right)^{-i}\mathbf{e}, \quad i \geq 1.$$

Some special cases of the continuous PH-distributions are,

1. The exponential distribution is a PH-distribution, where $\boldsymbol{\alpha} = 1$, $T = -\lambda$ and $m = 1$.

2. The Erlang distribution. Let $X_1, \ldots, X_n$ be independent with $X_i \sim exp(\lambda)$. Then $X_1 + \ldots + X_n$ is PH-distributed with $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$, and

$$T = \begin{pmatrix} -\lambda & \lambda & 0 & \ldots & 0 & 0 \\ 0 & -\lambda & \lambda & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & -\lambda \end{pmatrix}.$$

3. The Hypoexponential distribution or generalized Erlang distribution. Let $X_1, \ldots, X_m$ be independent with $X_i \sim exp(\lambda_i)$. Then $X_1 + \ldots + X_n$ is PH-distributed with $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$, and

$$T = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_m \end{pmatrix}.$$

4. The hyper-exponential distribution. Let $X_1, \ldots, X_m$ be independent with $X_i \sim exp(\lambda_i)$, and let f_i denote the corresponding exponential density. Let

$$f = \sum_{i=1}^{m} \alpha_i f_i, \quad \text{where } \alpha_i > 0, \sum_{i=1}^{m} \alpha_i = 1.$$

Then, the random variable that defines f is PH-distributed with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$, and

$$T = \begin{pmatrix} -\lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda_2 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_m \end{pmatrix}.$$

Continuous PH-distributions are a versatile and large class of distributions, with several important properties.

Theorem 1.1. *The class of PH-distributions is dense (in the sense of weak convergence) in the class off all positive-valued distributions.*

Theorem 1.1 implies that for every positive-valued distribution F, there is a sequence of PH-distributions which converges weakly to F. For a proof of Theorem 1.1, see Theorem 4.2 (page 84) of Asmussen [4]. Stochastic models that assume the exponential distribution have explicit solutions, and can be extended through the PH-distribution without loosing computational tractability. PH-distributions can also be defined with matrix representations and Markov chains, allowing a probabilistic interpretation, which introduces the use Markov models in the modeling framework, as we shall see later on. For further properties regarding the PH-distribution, we refer the reader to the works of Asmussen [4], Breuer and Baum [10], He [36] and Latouche and Ramaswami [54].

As commented previously, the assumption of independent and identically distributed exponential inter-event times is restrictive in practice. In this section the exponential distribution was generalized to PH-distributions. Next, the independence assumption of the inter-event times is extended by means of the Markov process.

1.2.2 Markov chains and Markov processes

Consider a stochastic process $\{X(t), t \in \mathcal{T}\}$, whose values or states are elements of a state space $\mathcal{S}$ and $\mathcal{T}$ denotes time. Then $\{X(t), t \geq 0\}$ is a Markov process if it satisfies the Markov property, which states: if for any $0 \leq t_0 \leq \ldots \leq t_k \leq t_{k+1}$ and $x_l \in \mathcal{S}$,

$$P[X(t_{k+1}) = x_{k+1}|X(t_k) = x_k, \ldots, X(t_0) = x_0] = P[X(t_{k+1}) = x_{k+1}|X(t_k) = x_k],$$

which means that given the present state of the process x_k at time t_k and all the previous states, the future state x_{k+1} depends only on the present state x_k and is independent from the past. That is, if we know the state of the stochastic process at a specific time, then we are able to predict future stochastic behavior. This makes the Markov processes helpful tools to stochastically model many real life problems. The Markov processes are classified according to the nature of the time parameter $\mathcal{T}$ and the nature of the state-space $\mathcal{S}$. In this dissertation we will deal with Markov processes with discrete state-space $\mathcal{S}$; and with a discrete and continuous parameter $\mathcal{T}$ which we define in the following subsections. We are considering Markov processes whose state of $\{X(t)\}$ does not depend on the time unit t, then the process $\{X(t)\}$ is said to have homogeneous transition probabilities.

1.2.2.1 Discrete time Markov Process

Consider a discrete-time stochastic process $\{X_n, n = 0, 1, 2, \ldots\}$, where $X_n = i$ denotes that the process is in state i at a discrete time n. Then $\{X_n, n \geq 0\}$ is a discrete time-homogeneous Markov Process, known as Markov chain, if the following holds

$$P[X_{n+1} = j|X_n = i, X_k = x_k, 0 \leq k \leq n-1] = P[X_{n+1} = j|X_n = i] = p_{ij},$$

for all non-negative integer $j, i, x_k, 0 \leq k \leq n-1$ and all $n \geq 0$. The previous property states that knowing the state of X_n, the future states of X_{n+1} do not depend on the previous states X_k for $0 \leq k \leq n-1$, but only on the present state. The state transition probability p_{ij} satisfies,

1. $p_{ij} \geq 0$,
2. $\sum_{j=0}^{\infty} p_{ij} = 1$, for $i = 0, 1, \ldots$.

The matrix P defined from the transition probabilities $\{p_{ij}\}$, is called the transition probability matrix, and is given by

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

The matrix P is a stochastic matrix, since for any row i, $\sum_{j=0}^{\infty} p_{ij} = 1$ holds. Let $p_{ij}^{(n)}$ denote the probability that the process goes into state j from state i after n transitions, that is

$$\begin{aligned} p_{ij}^{(n)} &= P[X_{m+n} = j | X_m = i], \\ p_{ij}^{(0)} &= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \\ p_{ij}^{(1)} &= p_{ij}. \end{aligned}$$

Then, the Chapman-Kolmogorov equations state that for all $n \geq 1$,

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

which in matrix form is, let $P^{(n)}$ denote the matrix of n-step transition probabilities $p_{ij}^{(n)}$. Then,

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

A Markov chain is fully characterized by the transition probability matrix P and an initial probability vector

$$\alpha_i = P[X_0 = i], \quad \text{where} \sum_{i=0}^{\infty} \alpha_i = 1.$$

Therefore, the probability that the process is in state j at the n-th transition, denoted by $\alpha_j^{(n)} = P[X_n = j]$, for all $n \geq 0$, is given by,

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

Let T be the time of first visit to state i. State i is called recurrent if $P(T < \infty | X_0 = i) = 1$; otherwise, if $P(T = +\infty | X_0 = i) > 0$,then i is called transient. A recurrent state i is called null recurrent if $E(T|X_0 = 1) = \infty$; otherwise, it is called positive recurrent. A recurrent state i is said to be periodic with period d if $d \geq 2$ is the largest integer for which $P(T = nd \text{ for some } n \geq 1 | X_0 = 1) = 1$; otherwise, if there is no such $d \geq 2$, i is called aperiodic.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 positive recurrent, or that all states are transient. Positive recurrent, aperiodic states are called ergodic. Finally, the next result presents the so-called stationary (or steady-state) probability vector.

Theorem 1.2. *For a finite, irreducible and aperiodic Markov chain, the limit*

$$\lim_{n \to \infty} P^n = \mathbf{e}\boldsymbol{\pi}$$

exists, where P is the transition probability matrix of the Markov chain. $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ is the unique stationary probability vector that satisfies

$$\boldsymbol{\pi} = \boldsymbol{\pi} P \qquad \textit{and} \qquad \boldsymbol{\pi}\mathbf{e} = 1,$$

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

This section dealt with Markov processes that evolved through a discrete time. The next section considers continuous-time Markov Processes, for which transitions can occur at any time along a continuous interval.

1.2.2.2 Continuous-time Markov Process

A continuous-time stochastic process $\{X(t), t \geq 0\}$ is a continuous-time Markov Process with discrete state-space, if the following holds

$$P[X(t+s) = j | X(s) = i, X(u) = u, 0 \leq u \leq s] = P[X(t+s) = j | X(s) = i] = p_{ij}(t)$$

for all $s, t \geq 0$ and non-negative integers $i, j, x(u), 0 \leq u \leq s$. In other words, knowing the state of $X(s)$, the future states of $X(s+t)$ does not depend on the previous states $X(u)$ for $0 \leq u \leq s$, and depends only on the present state. Let $p_{ij}(t)$ be the probability that the Markov process will be in state j, given that it departs from state i, after some additional time t. The quantity $p_{ij}(t)$ is called the transition probability function, and satisfies

1. $0 \leq p_{ij}(t) \leq 1$,
2. $\sum_j p_{ij}(t) = 1$,
3. $p_{ij}(t+s) = \sum_k p_{ik}(t)p_{kj}(s)$. This equation is the Chapman-Kolmogorov equation for the continuous-time Markov Process.

The matrix $P(t) = \{p_{ij}(t)\}$ is called the transition probability matrix at time t. Then, the Chapman-Kolmogorov equation can be rewritten as

$$P(t+s) = P(t)P(s).$$

When the process enters state i, it spends an amount of time in that state, called holding time or sojourn time, denoted by H_i. Then, due to the Markovian property, the following holds

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

So H_i is memoryless, hence H_i s exponentially distributed with mean $1/\lambda_i$. At the end of a sojourn time in state i, the process makes a transition to another state j with probability p_{ij}. Since the mean of H_i is $1/\lambda_i$, λ_i represents the rate at which the process leaves state i, and $\lambda_i p_{ij}$ denotes the rate when in state i, the process makes a transition into state j.

Define the following matrix $Q = \{q_{ij}\}_{i,j \in \mathcal{S}}$, where

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

Proposition 1.2. *Let $P(t)$ be the probability transition matrix of a continuous-time Markov process. Then the infinitesimal generator Q exists and is defined as,*

$$\lim_{t \to 0} \frac{P(t) - I}{t} = Q,$$

and moreover, for $t \geq 0$,

$$\frac{P(t)}{dt} = QP(t) = P(t)Q. \tag{1.1}$$

Remark 1.1. *The infinitesimal generator Q plays the same role as the transition matrix P of the discrete-time Markov chains. They are often also referred to as the rate matrix or intensity matrix of the Markov process.*

Since the state-space is assumed to be finite, then from the previous result, (1.1), it follows that

$$P(t) = e^{Qt}.$$

Every continuous-time Markov process has an associated embedded discrete-time Markov chain. If we consider the continuous-time Markov process $\{X(t), t \geq 0\}$ only at the instants upon which a state transition occurs, and we number these moments t_0, t_1, t_2,. . .., then we get a Markov chain $\{X_n, n \geq 0\}$, whose value is the state of $\{X(t)\}$ immediately after the transition at time t_n. The states of a Markov process can be classified by the classification provided by the embedded Markov chain.

Definition 1.4. *Let $\{X(t), t \geq 0\}$ be continuous-time Markov process, and let $\{X_n, n \geq 0\}$ be its associated embedded discrete-time Markov chain.*

1. *$\{X(t)\}$ is irreducible if and only if $\{X_n\}$ is irreducible.*

2. *A state i is recurrent/transient for $\{X(t)\}$ if and only if it is recurrent/transient for $\{X_n\}$.*

The following result establishes the limiting probability of an irreducible and positive recurrent Markov process.

Theorem 1.3. *Let $\{X(t), t \geq 0\}$ be an irreducible and positive recurrent Markov process, then there exists*

$$\pi_j = \lim_{t\to\infty} P[X(t) = j], \qquad \textit{for all } j$$

and it is independent of the initial state. $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ is the unique stationary probability vector that satisfies

$$\boldsymbol{\pi} Q = 0 \qquad \textit{and} \qquad \boldsymbol{\pi}\mathbf{e} = 1.$$

1.2.3 Markov Renewal Processes

A stochastic process that combines renewal processes and Markov chains is called a Markov Renewal process. Consider a stochastic process in which the transition from state to state occurs according to a Markov chain, and the time between two successive state transitions is a random variable, whose distribution is not always exponential and depends on the current state as well as the successive transition state.

Define a stochastic process $\{X_n, n \geq 0\}$, with state space $\mathcal{S}$, and a random variable S_n that denotes the time of the n-th event occurrence ($S_0 = 0$) at which the process transitions from one state to the other. X_n denotes the state of the process at time S_n.

Definition 1.5. *The bivariate stochastic process $\{(X_n, S_n), n \geq 0\}$ is called a Markov renewal process with state space $\mathcal{S}$ if*

$$\begin{aligned} P[X_{n+1} &= j, S_{n+1} - S_n \leq t | X_0, X_1, \ldots, X_n = i; S_0, S_1, \ldots, S_n] \\ &= P[X_{n+1} = j, S_{n+1} - S_n \leq t | X_n = i] \equiv K_{i,j}(t) \end{aligned}$$

for $n, t \geq 0$ and $i, j \in \mathcal{S}$.

The values $K_{i,j}(t)$ define the probability that the next arrival occurs within time t and that the next state is j (given it starts in state i). The matrix $K(t) = \{K_{i,j}(t)\}$ is known as the semi-Markov kernel of the Markov renewal process $\{X_n, S_n\}$. Define

$$P^{\star}_{i,j} = \lim_{t \to \infty} K_{i,j}(t) = P[X_{n+1} = j | X_n = i] \tag{1.2}$$

Proposition 1.3. *$\{X_n, n \geq 0\}$ is a Markov chain with state space $\mathcal{S}$ and transition probability matrix $P^{\star} = \{P^{\star}_{i,j}\}$ given by (1.2).*

Let $\{Y(t)\}$ with state space $\mathcal{S}$, denote the state of the process at time t, defined by

$$Y(t) = X_n, \quad S_n \leq t < S_{n+1}$$

is called a semi-Markov process. The Markov chain $\{X_n\}$ is said to be an embedded Markov chain, a process that governs transitions between states. Note that

$$\begin{aligned} K_{i,j}(t) &= P[X_{n+1} = j, S_{n+1} - S_n \leq t | X_n = i] \\ &= P[S_{n+1} - S_n \leq t | X_n = i, X_{n+1} = j] P[X_{n+1} = j | X_n = i] \end{aligned}$$

Since $P[X_{n+1} = j | X_n = i] = P^{\star}_{i,j}$, define $G_{i,j}(t)$ as the conditional probability that the time spent in state i given that the next transition is to state j is less than or equal to t, that is

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

If we assume that $P^\star_{i,j} = 0$ for some (i, j) then $K_{i,j}(t) = 0$, for all t, and $G_{i,j}(t) = 1$. Then, for $P^\star_{i,j} > 0$

$$G_{i,j}(t) = \frac{K_{i,j}(t)}{P^\star_{i,j}}.$$

This provides an important result regarding the structure of $\{X_n, T_n\}$.

Proposition 1.4. *For any integer $n \geq 1$ and $u_1, u_2, \ldots, u_n$ n real numbers, we have:*

$$P(T_1 - T_0 \leq u_1, \ldots, T_n - T_{n-1} \leq u_n | X_0, \ldots, X_n) = G_{X_0,X_1}(u_1)G_{X_1,X_2}(u_2)\ldots G_{X_{n-1},X_n}(u_n),$$

that is, the n random variables $T_1 - T_0, T_2 - T_1 \ldots, T_n - T_{n-1}$ are conditionally independent given $X_0, \ldots, X_n$.

For a more extensive literature on Markov Renewal processes, we refer the reader to Breuer and Baum [10], Ibe [81] and Nakagawa [73].

Now that the assumptions of independent and identically distributed exponential inter-event times have been relaxed by generalizing the exponential distributions with phase-type distributions and the independence among inter-event times via the Markov processes, a new model with non-exponential and dependent inter-event times is introduced next.

1.2.3.1 The Markovian Arrival Process

The Markovian Arrival Process (*MAP*) is a matrix generalization of the Poisson point process (defined in Section 1.2.1) which inherits the tractability of the Poisson model and, at the same time, extends its capabilities. They were introduced by Neuts [75] as a versatile Markovian point process, allowing for non-exponential and dependent event occurrence times. However, the current matrix description of a *MAP* arises in Lucantoni [61], providing a convenient and more tractable representation of the process. The *MAP*s are versatile point processes (the stationary *MAP* is dense in the class of point processes, see Asmussen and Koole [2]) which maintain the tractability of the Markovian structure and therefore they have been widely considered in a number of real-life contexts where dependent arrivals are commonly observed. For example, in reliability (Montoro-Cazorla et al. [67, 69, 71, 72]); in teletraffic (Kang [47], Casale et al. [14], Tseng and Wang [103]); insurance (Landriault and Shi [53], Li and Ren [60] and weather forecasting (Ramírez-Cobo et al. [89]). One of the most important applications of the *MAP* is in queuing theory, because it provides a way to model more complex arrival systems.

Consider a Poisson process with arrival rate λ. If $N(t)$ is the number of events in $(0,t]$, then $\{N(t)\}_{t\geq 0}$ is a Markov process on the state space $\{i : i \geq 0\}$ with infinitesimal generator of the form

$$Q_{POISSON} = \begin{pmatrix} d_0 & d_1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & d_0 & d_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & d_0 & d_1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & d_0 & d_1 & 0 & \cdots \\ \cdots & . & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

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

The m-state *MAP*, denoted MAP_m, is constructed by generalizing the above Poisson process to allow for non-exponential times between event occurrences, but still preserving an underlying Markovian structure. To accomplish this, consider a 2-dimensional Markov process $\{J(t), N(t)\}$ on the state space $\{(i,j) : i \geq 0; 1 \leq j \leq m\}$, with infinitesimal generator Q_{MAP} having the structure

$$Q_{MAP} = \begin{pmatrix} D_0 & D_1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & D_0 & D_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & D_0 & D_1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & D_0 & D_1 & 0 & \cdots \\ \cdots & . & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

where D_1 is a non-negative $m \times m$ matrix, D_0 has negative diagonal elements and non-negative off-diagonal elements. The process $J(t)$ represents an irreducible and continuous Markov process with state space $\mathcal{S} = \{1, \ldots, m\}$ and generator matrix D. The process $\{N(t), t \geq 0\}$ counts the number of events in the interval $(0,t]$. The MAP_m behaves as follows: the initial state $i_0 \in \mathcal{S}$ is generated according to an initial probability vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ 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}, there will be a transition from one state to another (necessarily different), and no event occur. On the other hand, with probability p_{ij1}, there will be a transition from one state to another (possibly the same), and an event occurs. The transition probabilities satisfy

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

When $m = 2$, we have a two-state *MAP*, denoted by MAP_2. Figure 1.8 illustrates the different transitions that can occur in this process by means of a transition diagram.

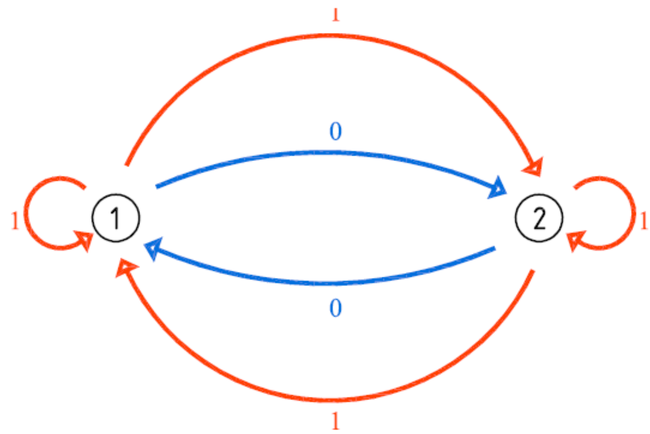

Figure 1.8: Transition diagram for the MAP_2. 0 and 1 illustrate moves without and with arrivals, respectively.

A MAP_m can thus be expressed in terms of $\{\boldsymbol{\alpha}, \boldsymbol{\lambda}, P_0, P_1\}$ where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and P_0, P_1 are $m \times m$ transition probability matrices with elements p_{ij0} $(i \neq j)$ and p_{ij1}, respectively. Instead of transition probability matrices, any MAP_m can be also characterized by $\mathcal{M} = \{\boldsymbol{\alpha}, D_0, D_1\}$, with elements given by

$$\begin{aligned} (D_0)_{ii} &= -\lambda_i, \quad i \in \mathcal{S}, \\ (D_0)_{ij} &= \lambda_i p_{ij0}, \quad i, j \in \mathcal{S},\ i \neq j, \\ (D_1)_{ij} &= \lambda_i p_{ij1}, \quad i, j \in \mathcal{S}, \end{aligned} \tag{1.3}$$

where D_0 and D_1 are $m \times m$ rate matrices governing the transitions without and with event occurrences, respectively. The definition of the rate matrices implies that $D = D_0 + D_1$ is the infinitesimal generator of the underlying Markov process $J(t)$. The matrix D_0 is assumed to be non-singular and the transition times are finite with probability 1. This implies that the point process does not terminate. The role of the states in this model is to provide inter-event times distributed as a random sum of non-identical exponentials. The stationary probability vector of the Markov process with generator D is $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$, which satisfies

$$\boldsymbol{\pi} D = \mathbf{0}, \quad \boldsymbol{\pi}\mathbf{e} = 1.$$

Thus, π_i represents the stationary probability that the process is in state i, for $i \in \mathcal{S}$. The constant λ^* is called the fundamental arrival rate and represents the expected number of event occurrences per unit of time of the MAP_m. It verifies

$$\lambda^* = \boldsymbol{\pi} D_1 \mathbf{e}$$

The *PH*-renewal processes and some non-renewal processes, as the Markov modulated Poisson process (*MMPP*) are well known special cases of the *MAP*. They are defined as follows:

1. A renewal process in which the inter-event times have a *PH*-distribution is called a *PH*-renewal process. The *PH*-renewal processes contains many familiar arrival processes including the Erlang and the hyperexponential arrival process. A *PH*-renewal process with representation $(\boldsymbol{\alpha}, T)$ is defined in an analogous way to the MAP_m, except that at an event occurrence, the new state of the Markov process is chosen according to the probability vector $\boldsymbol{\alpha}$, which is independent of the state from which an event occurred. In other words, a phase type renewal process is a MAP_m with $D_0 = T$ and $D_1 = -T\mathbf{e}\boldsymbol{\alpha}$. Therefore, in this case, the matrix D describes the Markov chain, obtained by resetting the original chain instantaneously using the same initial probabilities, whenever an event into the state occurs.

2. The Markov modulated Poisson process (*MMPP*) is a doubly stochastic Poisson process whose arrival rate is determined by a continuous time Markov process with a finite number of states, $\{J(t), t \geq 0\}$. The arrival rate therefore takes on only m values $\lambda_1, \ldots, \lambda_m$ and is equal to λ_j whenever $J(t) = j$. If the Markov process has infinitesimal generator R and if $\Lambda =$diag$(\lambda_1, \ldots, \lambda_m)$ then we have (in the *MAP* notation) $D_0 = R - \Lambda$ and $D_1 = \Lambda$. The Markov modulated Poisson process turns out to be a simplified *MAP* where arrivals only occur in transitions to the same state.

Several important properties of the MAP_m are reviewed in this Section. The first property concerns Markov renewal theory (see Chakravarthy [16]). Let X_n be the state of the underlying Markov process $J(t)$ at the time of the n-th event occurrence, and T_n the time between the $(n-1)$-th and the n-th occurrences, then $\{X_n, T_n\}_{n=1}^{\infty}$ is a Markov renewal process with semi-Markovian kernel given by

$$\int_0^t e^{D_0 t} D_1 dt = (I - e^{D_0 t})(-D_0)^{-1} D_1.$$

In particular, $\{X_n\}_{n=1}^{\infty}$ is a Markov chain whose transition probability matrix $P^{\star}$ is a irreducible stochastic matrix given by

$$P^{\star} = (-D_0)^{-1} D_1. \tag{1.4}$$

The stationary probability vector $\boldsymbol{\phi}$ of the Markov chain $\{X_n\}_{n=1}^{\infty}$ is defined as

$$\lim_{n \to \infty} \boldsymbol{\alpha}(P^{\star})^n = \boldsymbol{\phi},$$

which can be calculated solving the equation $\boldsymbol{\phi} P^{\star} = \boldsymbol{\phi}$.

If $\boldsymbol{\alpha} = \boldsymbol{\phi}$, where $\boldsymbol{\phi} = (\boldsymbol{\pi} D_1 \mathbf{e})^{-1} \boldsymbol{\pi} D_1$ (see Chakravarthy [16]), then the stationary version of the MAP_m is obtained. In what follows, we define the MAP_m descriptors for the non-stationary version. Let the random variable T_i denote the time between the $(i-1)$-th and i-th event occurrences in a non-stationary MAP_m. The sequence of random variables $\{T_i\}_{i \geq 1}$ are not identically distributed. They follow a PH-distribution with representation $\{\boldsymbol{\alpha}_i, D_0\}$ (see Chakravarthy [16] and Latouche and Ramaswami [54]), where

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} \left(P^{\star}\right)^{i-1},$$

with cumulative distribution given by

$$F_{T_i}(t) = 1 - \boldsymbol{\alpha}_i e^{D_0 t} \mathbf{e}., \quad \text{for } t \geq 0. \tag{1.5}$$

The moments of T_i can be computed as

$$\mu_{i,m} = E\left(T_i^m\right) = m! \boldsymbol{\alpha}_i \left(-D_0\right)^{-m} \mathbf{e}. \tag{1.6}$$

In addition, the auto-correlation function in the stationary version is given by

$$\rho(T_1, T_n) = \frac{\mu \boldsymbol{\pi} \left[(-D_0)^{-1} D_1\right]^{n-1} (-D_0)^{-1} \mathbf{e} - \mu^2}{\sigma^2},$$

where μ and σ^2 are the mean and variance of the process in its stationary version. Finally, the Laplace-Stieltjes transform of the n first consecutive inter-event times of a non-stationary MAP_m is given by

$$f^{*}_{T_1,\ldots,T_n}(s_1, \ldots, s_n) = \boldsymbol{\alpha}(s_1 I - D_0)^{-1} D_1 \ldots (s_n I - D_0)^{-1} D_1 \mathbf{e}. \tag{1.7}$$

The stationary counterparts of (1.5), (1.6) and (1.7), are easily derived by substituting $\boldsymbol{\alpha}_i$ or $\boldsymbol{\alpha}$, by $\boldsymbol{\phi}$, as the probability vector of the underlying Markov process $J(t)$. In the stationary version, the random variables $\{T_i\}_{i \geq 1}$ are identically distributed and $T \sim PH\left(\boldsymbol{\phi}, D_0\right)$.

Concerning the counting process $\{N(t),\ t \geq 0\}$, define $P(n,t) = \{P_{ij}(n,t)\}_{n \in \mathbb{N}, t \geq 0}$, as the $m \times m$ matrices whose (i,j)-th element is given by

$$P_{ij}(n,t) = P\left(N(t) = n,\ J(t) = j \mid N(0) = 0,\ J(0) = i\right), \tag{1.8}$$

for $1 \leq i, j \leq 2$. That is, the matrix $P(n,t)$ 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 time 0 there have been no events and the state is i. Then,

$$P\left(N(t) = n \mid N(0) = 0\right) = \boldsymbol{\alpha} P(n,t) \mathbf{e}, \tag{1.9}$$

and the expected number of event occurrences at time t, $E\left(N(t) \mid N(0)=0\right)$, is computed from the first factorial moment of the counting process,

$$M_1(t) = \sum_{n=0}^{\infty} nP(n,t), \tag{1.10}$$

where more details can be found in Chakravarthy [16] and Neuts and Li [77].

1.2.3.2 The Batch Markovian Arrival Process

In many contexts, it is not uncommon that the events occur in groups, that is, several simultaneous events may occur at the same time. In this setting, Lucantoni [62] proposed the Batch Markovian Arrival Process (*BMAP*), which is a generalization of the *MAP* that allows for correlated batch event occurrences. As with the *MAP*, the stationary *BMAP*s are capable of approximating any stationary batch point process (see Asmussen and Koole [2]), which suggests the versatility and range of applications of such processes. For a recent account of the literature on *BMAP*s applications, we refer the reader to Bookbinder et al. [8]; Falin [28]; Gómez-Corral and Economou [35]; Heckmüller and Wolfinger [40]; Kim et al. [48]; Kim and Kim [49]; Kim et al. [50]; Klemm et al. [51]; Niyato et al. [78].

To motivate the *BMAP*, consider a Poisson process with batch (group) arrivals. Let the rate of the Poisson process be λ and the probability that the batch size equals k be p_k, $k \geq 1$. If $N(t)$ is the number of arrivals in $(0,t]$, then $\{N(t)\}_{t\geq 0}$ is then a Markov process on the state space $\{i : i \geq 0\}$ with infinitesimal generator of the form

$$Q_{B-POISSON} = \begin{pmatrix} d_0 & d_1 & d_2 & d_3 & \cdots & \cdots \\ 0 & d_0 & d_1 & d_2 & \cdots & \cdots \\ 0 & 0 & d_0 & d_1 & \cdots & \cdots \\ 0 & 0 & 0 & d_0 & \cdots & \cdots \\ \cdots & . & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where, $d_0 = -\lambda$ and $d_k = \lambda p_k$ for $k \geq 1$. After an exponential sojourn (with mean λ^{-1}) in state i, the process jumps to state $i+k$ with probability p_k where the transition corresponds to an arrival and k corresponds to the size of the batch.

The m-state *BMAP* with maximum batch arrival size k, defined as $BMAP_m(k)$, is constructed by generalizing the above batch Poisson process to allow for non-exponential times between the arrivals of batches, but still preserving an underlying Markovian structure. To accomplish this, it is considered a two-dimensional Markov process $\{N(t), J(t)\}$ on the state space $\{(i,j) : i \geq 0, 1 \leq j \leq m\}$ with an infinitesimal generator Q_{BMAP} having the structure

$$Q_{BMAP} = \begin{pmatrix} D_0 & D_1 & D_2 & D_3 & \cdots \\ 0 & D_0 & D_1 & D_2 & \cdots \\ 0 & 0 & D_0 & D_1 & \cdots \\ 0 & 0 & 0 & D_0 & \cdots \\ \cdots & . & \cdots & \cdots & \cdots \end{pmatrix},$$

where D_k, $k \geq 0$ are $m \times m$ matrices, D_0 has negative diagonal elements and non-negative off-diagonal elements, D_k, $k \geq 1$, are non-negative, and D_g, defined by

$$D_g = \sum_{k=0}^{\infty} D_k$$

is an irreducible infinitesimal generator. Analogous to the MAP_m, 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}, $j \in \mathcal{S}$, no arrival occurs and the $BMAP_m(k)$ enters a different state $j \neq i$. On the other hand, with probability p_{ijl}, $1 \leq l \leq k$, $j \in \mathcal{S}$, there will be a transition to state j with a batch arrival of size l. The transition probabilities satisfy

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

Figure 1.9 illustrates by means of a transition diagram the different transitions that can occur in the $BMAP_2(2)$. The values 0, 1 and 2 correspond to transitions with no arrival, a single arrival or two arrivals, respectively. As the MAP_m, the $BMAP_m(k)$ can be expressed in terms of $\{\boldsymbol{\lambda}, P_0, P_1, \ldots, P_k\}$ where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $P_0, \ldots, P_k$ are $m \times m$ transition probability matrices with elements p_{ij0} $(i \neq j)$, p_{ij1},...,p_{ijk}, respectively. Instead of transition probability matrices, any $BMAP_m(k)$ can be also characterized by the rate matrices $\mathcal{B} = \{D_0, D_1, \ldots, D_k\}$ with elements given by

$$\begin{aligned} (D_0)_{ii} &= -\lambda_i, \quad i \in \mathcal{S}, \\ (D_0)_{ij} &= \lambda_i p_{ij0}, \quad i,j \in \mathcal{S},\ i \neq j, \\ (D_l)_{ij} &= \lambda_i p_{ijl}, \quad i,j \in \mathcal{S},\ 1 \leq l \leq k. \end{aligned} \tag{1.11}$$

The infinitesimal generator of the underlying Markov process $J(t)$ is given by,

$$Q = \sum_{l=0}^{k} D_l, \tag{1.12}$$

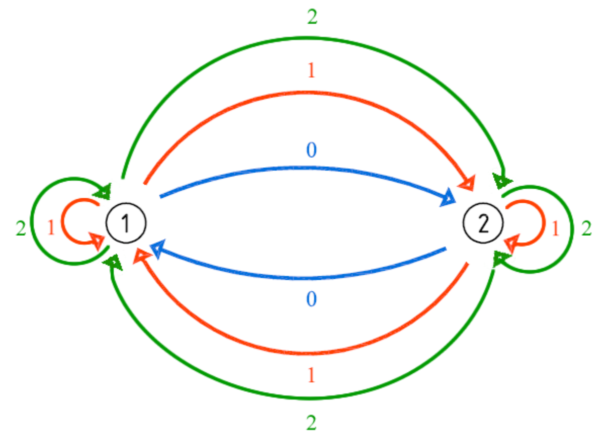

Figure 1.9: Transition diagram for the $BMAP_2(2)$. 0 denotes moves without arrivals and 1 and 2 denotes moves with respective batch arrivals.

with stationary probability vector $\boldsymbol{\pi}_{BMAP} = (\pi_{BMAP}(1), \ldots, \pi_{BMAP}(m))$, such that

$$\boldsymbol{\pi}_{BMAP} Q = \mathbf{0}, \quad \boldsymbol{\pi}_{BMAP}\mathbf{e} = 1.$$

The fundamental arrival rate for the $BMAP$ is

$$\lambda^{\star}_{BMAP} = \boldsymbol{\pi}_{BMAP} \sum_{l=1}^{k} l D_l \mathbf{e}.$$

The $BMAP$ includes well-known families of processes such as the Batch PH-renewal processes and the Batch Markov Modulated Poisson process. Recall the definitions of PH-renewal processes and the Markov Modulated Poisson process given in the previous subsection, then their batch counterparts may be defined as follows:

1. Batch PH-renewal processes. Let p_k denote the probability that the batch size is k, $k \geq 1$. Then the batch PH-renewal processes is a $BMAP_m(k)$ with $D_0 = T$ and $D_k = p_k T^0 \boldsymbol{\alpha}$, $k \geq 1$, where $T^0 = -T\mathbf{e}$.

2. Batch Markov Modulated Poisson process. Let p_k denote the probability that the batch size is k, $k \geq 1$. Then the batch Markov Modulated Poisson process is a $BMAP_m(k)$ with $D_0 = R - \Lambda$, and $D_k = p_k \Lambda$ for $k \geq 1$.

The properties of the $BMAP_m(k)$ are analogous to the ones given for the MAP_m. Let

$$D = \sum_{l=1}^{k} D_l,$$

then, the transition probability matrix $P^\star$ is a irreducible stochastic matrix given by

$$P^\star = (-D_0)^{-1}D, \tag{1.13}$$

whose stationary probability vector $\boldsymbol{\phi}_{BMAP}$ is computed as,

$$\boldsymbol{\phi}_{BMAP} = (\boldsymbol{\pi}_{BMAP}D\mathbf{e})^{-1}\boldsymbol{\pi}_{BMAP}D, \tag{1.14}$$

(see Ramírez-Cobo et al. [87] for a proof).

Let the random variable T_i, $i \geq 1$, denote the time between the $(i-1)$-th and i-th event occurrences in a non-stationary $BMAP_m(k)$, $m, k \geq 2$. As in the MAP_m, the sequence of random variables $\{T_i\}_{i\geq 1}$ are not identically distributed, but follow a PH-distribution with representation $\{\boldsymbol{\alpha}_i, D_0\}$. Then the cumulative distribution and the moments of T_i are computed as in (1.5) and (1.6), respectively. In addition, the auto-correlation function of the inter-event times in the stationary version of the process is given by

$$\rho(T_1, T_n) = \frac{\mu\boldsymbol{\pi}_{BMAP}\left[(-D_0)^{-1}D\right]^{n-1}(-D_0)^{-1}\mathbf{e} - \mu^2}{\sigma^2},$$

where μ and σ^2 are the mean and variance of the inter-event times. The Laplace transform of the n first consecutive inter-event times of a non-stationary $BMAP_m(k)$, $m, k \geq 2$, is given by

$$f^*_{T_1,\ldots,T_n}(s_1, \ldots, s_n) = \boldsymbol{\alpha}(s_1 I - D_0)^{-1}D \ldots (s_n I - D_0)^{-1}D\mathbf{e}.$$

Finally, the counting process $\{N(t), t \geq 0\}$ is defined analogously as in the case of the MAP (see page 20).

For a thorough definition of the general $m-$state $BMAP$ we refer the reader to Chakravarthy [15], Cordeiro and Kharoufeh [20] and Lucantoni [62, 63]. It is important to point out the usefulness of the MAPs and $BMAP$s in queueing theory to model either the arrival process or the service process. The literature on queueing systems using the MAP is quite extensive. A detailed theoretical analysis of the single-server queue where the arrival process is governed by a $BMAP$ was first considered by Ramaswami [86]. Basic results for the steady-state analysis of the $MAP/G/1$ queue are provided in Lucantoni [61]. Several variants of the $MAP/G/1$ queue have been considered, for example, Gómez-Corral [34] and Li and Zhao [58]. The $BMAP/G/1$ has been studied by Lucantoni [62, 63], where new results were obtained and numerical algorithms leading to the computation of steady-state solutions, were derived. Several variants of the $BMAP/G/1$ queue have been considered, for example, see Ferng and Chang [29] and Li et al. [59].

1.3 Statistical estimation of Point processes

In the previous sections realistic probabilistic models for inter-event time data were described. In this section a review of the statistical estimation for such models is given.

As we have pointed out, point processes have the ability to describe a wide range of situations, therefore they have gained widespread use in stochastic modelling. Therefore, it is of interest to consider statistical inference for such models, a problem that has been less addressed in the literature than more theoretically-based questions. Let us recall from Section 1.2 that a point process can be specified by the event times, $\{S_n, n \geq 1\}$, the inter-event times, $\{T_n, n \geq 1\}$ or by the counting process $\{N(t), t \geq 0\}$, hence, the observations of any of such random processes will constitute the sample data for the statistical inference. The formalization of the problem is stated next. Consider a sample of length n, $\mathcal{X} = (X_1, \ldots, X_n)$ from a random variable $\mathcal{X}$ following a given probability model defined in terms of the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. The aim is to find point estimates of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, such that the model $\mathcal{X}$ under $\hat{\boldsymbol{\theta}}$ better fits the data. The most common techniques used in estimating general point processes are the following,

1. Method of Moments. In this case, the estimates are the solution to the following system of equations given by matching the theoretical moments with the sample moments. Let us recall the j-th theoretical moment of a random variable is given by,

$$\mu_j \equiv \mu_j(\boldsymbol{\theta}) = E\left(X^j\right), \tag{1.15}$$

and the j-th sample moment is,

$$\hat{\mu_j} = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \tag{1.16}$$

then the parameter estimates, $\hat{\boldsymbol{\theta}}$ is defined to be value of $\boldsymbol{\theta}$ that solve the system of k equations given by,

$$\mu_i(\hat{\boldsymbol{\theta}}) = \hat{\mu}_i, \qquad 1 \leq i \leq k.$$

This method is also useful to obtain starting points for approaches that require iterative numerical routines [105].

2. Least-squares. It is based on minimizing the squared discrepancies between observed data and their expected values. Formally, consider the theoretical and sample moments

defined in (1.15) and (1.16), then the objective is to find the parameter estimates, $\hat{\theta}$, solving the following optimization problem

$$\min_{\hat{\boldsymbol{\theta}}} \left(\sum_{i \in \mathcal{M}} \left(\beta_i \frac{\mu_i}{\hat{\mu}_i} - \beta_i \right)^2 \right),$$

where $\mathcal{M}$ is a set of moments to be approximated and β_i are optional weights.

3. Maximum likelihood. Suppose that $\mathcal{X}$ with density function f is observed, and define the likelihood function by,

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{X}) = f(X_1, X_2, \ldots, X_n|\boldsymbol{\theta}),$$

and the log-likelihood function by $\log(\mathcal{L}(\boldsymbol{\theta}|\mathcal{X})) = \log(f(X_1, X_2, \ldots, X_n|\boldsymbol{\theta}))$. Then the maximum likelihood approach is to find a set of parameter estimates, $\hat{\boldsymbol{\theta}}$, that maximizes the likelihood $f(\mathcal{X}|\boldsymbol{\theta})$, that is, solve the following optimization problem,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathcal{X})$$

4. Expectation-Maximization (EM) approach. The EM algorithm (see Dempster [23]) is an iterative method useful in problems with incomplete data. Each iteration of the algorithm works in two steps, an expectation (E) and a maximization (M) step.

   Suppose we have incomplete observed data $\mathcal{X}$, and the whole data is given by the set $(\mathcal{X}, \mathcal{T})$, where $\mathcal{T}$ is a set of unobserved latent data. The algorithm is as follows: first pick a starting value $\boldsymbol{\theta}_0$. Now, for $j \geq 1$ repeat steps 1 and 2 below until convergence.

   (1) (E-step) Calculate
   $$J(\boldsymbol{\theta}|\boldsymbol{\theta}^j) = E\left(\log f(\mathcal{X}, \mathcal{T}|\boldsymbol{\theta})|\mathcal{X}, \boldsymbol{\theta}^j\right)$$

   (2) (M-step) Find the parameter that maximizes
   $$\boldsymbol{\theta}^{j+1} = \arg\max_{\boldsymbol{\theta}} J(\boldsymbol{\theta}|\boldsymbol{\theta}^j)$$

   The E-step consists in finding the distribution of the unobserved data, given the observed data set and the current value of the parameters $\boldsymbol{\theta}^j$, so that the M-step reestimates the parameters with a maximum-likelihood approach.

We next review the statistical estimation for the point processes this dissertation deals with, the *(B)MAP*. First, we have to point out that there are several issues to take into account when studying the estimation of these processes. These processes are complex models that include transitions to states between event occurrences. However, in practice, the only observable information are the inter-event times and the batch sizes, the underlying Markov chain transitions are not available. Processes where the underlying Markov process is not observable, but the sequence of observations is available, are called hidden Markov models. Therefore, in the case of the *(B)MAP*, the observed data can be viewed as being generated from a hidden Markov process, see [26]. The lack of identifiability is very common when dealing with hidden Markov processes, and implies that different parametrizations represent the same process; that is, they have the same likelihood function for any sequence of inter-event times. Therefore, in the context of statistical inference, it is important to obtain a unique parametrization of the model. Identifiability conditions for general hidden Markov processes are studied in Backwell and Koopmans [6], Ito et al. [46], Leroux [56] and Rydén [93]. In the literature, there are many works dedicated to the identifiability issue for some cases of the *MAP*. It is well known that phase-type representations are not unique (see O´ Cinneide [79]), however, since any two-state *PH*-distribution can be transformed into an acyclic form *PH*-distribution (see [21]), and the canonical form is provided for the acyclic phase type distributions (see [22]) of any order, it is shown that the two-state *PH*-distribution has a canonical form. Also the three-state case canonical form has been provided by Horváth and Telek [44], the case for $m \geq 4$ is still an open research problem. It was shown by Rydén [94] that the Markov modulated Poisson process (*MMPP*) can be identified up to permutations of states. Ramírez-Cobo et al. [87] provided the conditions for which stationary MAP_2s are equivalent, and Bodrog et al. [7] provided a canonical and unique representation of the stationary MAP_2. An extensive description of identifiability will be given in Chapter 2 and Chapter 4.

A moment matching approach for estimating the *MAP* has been studied in Horváth and Telek [43], Telek and Horváth [102], Eum et al. [27], Bodrog et al. [7], Casale et al [14], but in these works the identifiability of the model was not considered (except in [102, 7]). The EM algorithm has been proposed for inference of the *MAP* (see Asmussen et al [3], Rydén [95], Klemm et al. [51] and Buchholz [11]) and for the *BMAP* (Breuer [9] and Okamura et al. [80]), since it is a general method for computing maximum likelihood estimations in statistical models in which there exist random variables which are not observable. A maximum likelihood approach was tackled by Carrizosa and Ramírez [12] for the *MAP*. A number of Bayesian approaches for estimating the *MMPP* can be found in the literature: Scott [100] developed a Bayesian inference for the two-state *MMPP* and Fearnhead and Sherlock [30] derived a Gibbs sampler that samples from the exact distribution of the hidden

Markov chain in a *MMPP*. Bayesian inference for the MAP_2 has been studied by Ramírez-Cobo et al. [90], where different algorithms are proposed (Metropoling-Hastings within Gibbs sampling) in which the unobserved data are partially reconstructed or not reconstructed at all, in order to approximate the inter-event time distributions and estimate some identifiable quantities.

1.4 Structure of this dissertation

This dissertation is composed of 6 Chapters. Chapter 1 presented the background that introduces the most important concepts and properties that are needed to carry out our analyzes. After providing a brief review of point processes and renewal processes, the generalization of the classical Poisson process assumptions to the phase-type distributions and Markov processes was given. It was followed by a description of the *MAP* and *BMAP* and its main properties. The chapter concludes with a review of the estimation procedures of point processes and *(B)MAP*s that have been proposed in the literature.

The theoretical contributions of this thesis are developed in Chapters 2, 3, 4 and 5. The results in Chapter 2 concern the identifiability of the non-stationary two-state *MAP*. It is proven that the usual parametrization of the process is not unique, which means that the process is nonidentifiable given the inter-event times. We propose a method to construct non-stationary two-state *MAP*s from any fixed one. Finally, a canonical and unique representation of the non-stationary two-state *MAP* is provided.

In Chapter 3, motivated by the unique canonical representation found in the previous chapter, Rodríguez et al. [98] studied an estimation approach for the non-stationary two state *MAP*. The data to be considered are the failures of N electrical components that are assumed to be identically distributed, but for which it is not reasonable to assume that the inter-failure times related to each component are independent nor identically distributed. A moment matching estimation approach is proposed to fit the data via a non-stationary two-state *MAP*. We also provide a simulated and a real data set provided by the Spanish electrical group Iberdrola to illustrate our approach.

With the aim to extend the properties of *MAP*s to the *BMAP*, which introduces correlated arrivals in batches, we consider the identifiability issue of the *BMAP* in Chapter 4 (see Rodríguez et al. [97]). This chapter investigates the identifiability issue of the stationary two-state *BMAP* noted as $BMAP_2(k)$, where k is the maximum batch arrival size, under the assumptions that both the inter-event times and batches sizes are observed. We prove that for $k \geq 2$ the process cannot be identified. The proof is based on the construction of an equivalent $BMAP_2(k)$ to a given one, and on the decomposition of a $BMAP_2(k)$ into k $BMAP_2(2)$s. We illustrate our findings with numerical examples.

The auto-correlation functions of the inter-event times and batch sizes of the $BMAP$ are studied in Chapter 5. This chapter examines the characterization of both auto-correlation functions for the stationary $BMAP_2(k)$, for $k \geq 2$, patterns of behavior are identified for both cases for the $BMAP_2(2)$. It is proven that both auto-correlation functions decrease geometrically as the time lag increases. Also, the characterization of the auto-correlation functions has been extended for the general $BMAP_m(k)$ case, $m \geq 3$.

Finally, in Chapter 6 we summarize the most significant contributions of this dissertation, and also provide a short description of possible research lines.

Chapter 2

A canonical form for the non-stationary MAP_2

The Markovian Arrival Process (*MAP*) was introduced in Section 1.2.3.1 as a generalization of the Poisson point process by allowing non-exponential and dependent inter-event times. This makes the *MAP* a versatile and flexible option to model non-Poisson real data. The *MAP* has been considered in several applied problems, for instance, in reliability, queueing or teletraffic, we refer the reader to the works of Montoro-Cazorla et al. [67, 69], Okamura et al. [80], Kang et al. [47], Casale et al. [14], Wu et al [106] and Ramírez-Cobo et al. [88, 89].

Real-life data observations usually correspond to inter-event times and both state transitions and transitions where none event occurs remain unobserved. In this sense, data can be viewed as being generated from a hidden Markov process, which is commonly characterized by a lack of identifiability. The identifiability problem was introduced in Section 1.3, and it occurs when the parameters of a model are not uniquely determined, in the sense that the likelihood function is the same under at least two model representations. The identifiability problem for general hidden Markov processes are studied in Blackwell and Koopmans [6], Ito et al. [46], Leroux [56] and Rydén [93]. A few works have studied the identifiability issue for some cases of *MAP*s. For example, the phase-type distributions lack a unique representation, see O´ Cinneide [79]. Also, He and Zhang [37, 38, 39] study the identifiability problem of Coxian distribution, which is a phase-type related distribution. In contrast, Rydén [94] showed that the Markov modulated Poisson process (*MMPP*) can be identified up to permutations of states. To the best of our knowledge, identifiability for the non-stationary *MAP* has not been considered before. However, Telek and Horváth [102] investigated a minimal representation for the stationary m-state MAP, and Bodrog et al. [7] provided a canonical and unique representation for the stationary two-state MAP. Ramírez-Cobo et al. [87] provided a procedure to build infinite equivalent stationary two-state MAP for a given fixed one, which states the conditions that two representations need to have in order to be equivalent. Also Ramírez-Cobo and Lillo [91] partially solved the problem for the stationary three-state MAP. Unlike the stationary version of the process, in the non-stationary MAP the inter-event times are not identically distributed. This fact makes the non-stationary MAP to have more applicability in terms of modeling than its stationary counterpart. In order to develop an estimation method to fit the model to real data sets, a detailed examination of certain properties of the process is needed. In particular, the identifiability of the process is crucial, which, as previously commented, determines the possible multimodality of the likelihood function.

In this chapter, we aim to contribute a detailed study of the identifiability problem for the non-stationary two-state MAP (MAP_2). In Section 2.1, we prove that the non-stationary MAP_2 is a non-identifiable process, we establish a methodology to construct equivalent non-stationary MAP_2s. In Section 2.2, following Bodrog et al. [7], we present a canonical, unique representation of the process. In Section 2.3 we provide a moments based characterization of the process. Finally, in Section 2.4 we give the conclusions.

2.1 Equivalent non-stationary MAP_2

From now on, a non-stationary MAP_2 will be represented by $\mathcal{M} = \{\boldsymbol{\alpha}, D_0, D_1\}$ where

$$\boldsymbol{\alpha} = (\alpha,\ 1-\alpha), \quad D_0 = \begin{pmatrix} x & y \\ z & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} w & -x-y-w \\ v & -z-u-v \end{pmatrix}, \tag{2.1}$$

and without loss of generality it is assumed that $u \leq x$. According to (1.3)

$$x = -\lambda_1, \quad y = \lambda_1 p_{120}, \quad w = \lambda_1 p_{111},$$
$$z = \lambda_2 p_{210}, \quad u = -\lambda_2, \quad v = \lambda_2 p_{211}.$$

Let $\boldsymbol{\pi} = (\pi, 1-\pi)$ be the stationary probability of the underlying Markov process $J(t)$ of the non-stationary MAP_2. If $\boldsymbol{\alpha} = \boldsymbol{\phi}$, where $\boldsymbol{\phi}$ is the stationary probability vector of the Markov chain related to transitions with event occurrences ($\boldsymbol{\phi} = (\boldsymbol{\pi} D_1 \mathbf{e})^{-1} \boldsymbol{\pi} D_1$, see Chakravarthy [15]), then we have the stationary version of the MAP_2.

In practice, it is usual that only the inter-event times are observed when inference of the process is considered. Therefore, our definition of identifiability is stated in the way specified by Rydén [94] and Ramírez-Cobo et al. [87]; that is, in terms of the inter-event times distribution.

Definition 2.1. *The non-stationary MAP_2 is a non-identifiable process if for any fixed MAP_2 with representation $\mathcal{M}$, then there exists another MAP_2 with different representation $\widetilde{\mathcal{M}}$ such that*

$$\left(T_1, \ldots, T_n\right) \stackrel{d}{=} \left(\widetilde{T}_1, \ldots, \widetilde{T}_n\right), \quad \textit{for all } n \geq 1, \tag{2.2}$$

where T_i represents the time between the $(i-1)$-th and i-th event occurrences in the MAP_2 defined by $\mathcal{M}$ (similarly define $\widetilde{T}_i$).

Note that the equality in distribution (2.2) is equivalent to the equality of the Laplace transforms defined in (1.7),

$$f^*_{T;\boldsymbol{\alpha},D_0,D_1}\left(s_1, \ldots, s_n\right) = f^*_{\tilde{T};\tilde{\boldsymbol{\alpha}},\tilde{D}_0,\tilde{D}_1}\left(s_1, \ldots, s_n\right), \quad \text{for all } n \geq 1. \tag{2.3}$$

We state that two MAP_2s are equivalent if they posses the same Laplace transforms, that is if (2.3) is satisfied.

The non-identifiability of the non-stationary MAP_m is to be expected, since an important subset of the non-stationary process has been proven to be non-identifiable, which is the stationary MAP_m, see Telek and Horváth [102]. However, our main interest resides in finding the canonical form of the non-stationary MAP_2 and the inference of its parameters. In order to achieve this goal, we will examine in detail the identifiability issue of the process.

In the next result, we prove the non-identifiability of the non-stationary MAP_m by means of a similarity transform, which will be a useful tool in the following sections of this chapter.

Theorem 2.1. *Consider two non-stationary MAP_ms, $\mathcal{M}$ and $\widetilde{\mathcal{M}}$. Then (2.3) holds if and only if there exists an invertible matrix B whose row sums equal to one such that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$, $\tilde{D}_0 = B D_0 B^{-1}$ and $\tilde{D}_1 = B D_1 B^{-1}$.*

Proof. Consider an arbitrary invertible matrix B such that $B\mathbf{e} = \mathbf{e}$. Then through a similarity transform with B, a MAP_m defined by the parameters $\{\tilde{\boldsymbol{\alpha}}, \tilde{D}_0, \tilde{D}_1\}$, where $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$, $\tilde{D}_0 = B D_0 B^{-1}$ and $\tilde{D}_1 = B D_1 B^{-1}$, is equivalent to that defined by $\{\boldsymbol{\alpha}, D_0, D_1\}$. That is,

$$\begin{aligned} f^*_{T;\boldsymbol{\alpha},D_0,D_1}\left(s_1, \ldots, s_n\right) &= \boldsymbol{\alpha} B^{-1} B (s_1 I - D_0)^{-1} B^{-1} B D_1 B^{-1} \\ &\times \; B (s_2 I - D_0)^{-1} B^{-1} B D_1 B^{-1} \\ &\times \; \ldots B (s_n I - D_0)^{-1} B^{-1} B D_1 B^{-1} B \mathbf{e}. \end{aligned}$$

Regrouping the factors in the above expression,

$$\begin{aligned} f^*_{T;\boldsymbol{\alpha},D_0,D_1}(s_1,\ldots,s_n) &= \boldsymbol{\alpha}B^{-1}(s_1I-BD_0B^{-1})^{-1}BD_1B^{-1} \\ &\times\ (s_2I-BD_0B^{-1})^{-1}BD_1B^{-1} \\ &\times\ \ldots(s_nI-BD_0B^{-1})^{-1}BD_1B^{-1}B\mathbf{e} \\ &= f^*_{\tilde{T};\tilde{\boldsymbol{\alpha}},\tilde{D}_0,\tilde{D}_1}(s_1,\ldots,s_n)\,. \end{aligned}$$

Assume that (2.3) holds. We know that the non-stationary MAP_2 converges to its stationary version. Telek and Horváth [102] (Theorem 3, page 1157) proved for the stationary version that if (2.3) holds, then there exists an invertible matrix B such that $\tilde{D}_0 = BD_0B^{-1}$, $\tilde{D}_1 = BD_1B^{-1}$ and $B\mathbf{e} = \mathbf{e}$, then this affirmation will also hold for the rate matrices in the non-stationary version. It remains to prove that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}B^{-1}$. If $\tilde{D}_0 = BD_0B^{-1}$ and $\tilde{D}_1 = BD_1B^{-1}$, then $\tilde{P}^\star = \left(-\tilde{D}_0\right)^{-1}\tilde{D}_1 = BP^\star B^{-1}$.

Since (2.3) is satisfied, then the equality of the moments of the inter-event time occurrences holds, that is, $E\left(T_i^m\right) = E\left(\tilde{T}_i^m\right)$, or equivalently, from (1.6)

$$\boldsymbol{\alpha}_i\left(-D_0\right)^{-m}\mathbf{e} = \tilde{\boldsymbol{\alpha}}_i\left(-\tilde{D}_0\right)^{-m}\mathbf{e},\ \text{for all } i,m\geq 1,$$

where $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}\left(P^\star\right)^{i-1}$ and $\tilde{\boldsymbol{\alpha}}_i = \tilde{\boldsymbol{\alpha}}\left(\tilde{P}^\star\right)^{i-1}$. Substituting $\tilde{D}_0$ and $\tilde{D}_1$ into the previous expression,

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}\left(\tilde{P}^\star\right)^{i-1}\left(-\tilde{D}_0\right)^{-m}\mathbf{e} &= \tilde{\boldsymbol{\alpha}}\left(BP^\star B^{-1}\right)^{i-1}\left(B(-D_0)B^{-1}\right)^{-m}\mathbf{e} \\ &= \tilde{\boldsymbol{\alpha}}B\left(P^\star\right)^{i-1}B^{-1}B\left(-D_0\right)^{-m}B^{-1}\mathbf{e} \\ &= \tilde{\boldsymbol{\alpha}}B\left(P^\star\right)^{i-1}\left(-D_0\right)^{-m}\mathbf{e}, \end{aligned}$$

therefore

$$\boldsymbol{\alpha}\left(P^\star\right)^{i-1}\left(-D_0\right)^{-m}\mathbf{e} = \tilde{\boldsymbol{\alpha}}B\left(P^\star\right)^{i-1}\left(-D_0\right)^{-m}\mathbf{e},\ \text{for all } i,m\geq 1. \tag{2.4}$$

Assume now that the spectral decomposition of $\left(-D_0\right)^{-1}$ and $P^\star$ are given by

$$\left(-D_0\right)^{-1} = VTV^{-1} \text{ and } P^\star = USU^{-1}.$$

Then, substituting the spectral decompositions into (2.4) leads to

$$\begin{aligned} \boldsymbol{\alpha}\left(P^\star\right)^{i-1}\left(-D_0\right)^{-m}\mathbf{e} &= \boldsymbol{\alpha}\left(USU^{-1}\right)^{i-1}\left(VTV^{-1}\right)^{m}\mathbf{e} \\ &= \boldsymbol{\alpha}US^{i-1}U^{-1}VT^mV^{-1}\mathbf{e} \\ &= \tilde{\boldsymbol{\alpha}}BUS^{i-1}U^{-1}VT^mV^{-1}\mathbf{e},\ \text{for all } i,m\geq 1. \end{aligned}$$

Due to the fact that $\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\alpha}}$, B and $\mathbf{e}$ are nonzero and D_0 is an invertible matrix, then (2.4) implies $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}B$, and therefore $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}B^{-1}$.

□

Now that the non-identifiability of the MAP_m has been proven, we proceed to study in detail the consequences of such lack of identifiability in the case of $m = 2$. The most natural approach to construct equivalent non-stationary MAP_2s for any given one is to choose an unknown invertible matrix B that satisfies $B\mathbf{e} = \mathbf{e}$ and carry out a similarity transform. But this technique may not generate a real non-stationary MAP_2, in the sense that the parameters of the process are not well-defined.

Example 2.1. *Consider the non-stationary MAP_2 given by*

$$\boldsymbol{\alpha} = (0.0220, 0.9780), \quad D_0 = \begin{pmatrix} -5.5123 & 5.5019 \\ 2.8913 & -3.6643 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0009 & 0.0095 \\ 0.1105 & 0.6625 \end{pmatrix},$$

and the arbitrary invertible, stochastic matrix

$$B = \begin{pmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{pmatrix}.$$

From $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}B^{-1}$, $\tilde{D}_0 = BD_0B^{-1}$ and $\tilde{D}_1 = BD_1B^{-1}$,

$$\tilde{\boldsymbol{\alpha}} = (3.7805, -2.7805), \quad \tilde{D}_0 = \begin{pmatrix} -5.8791 & 5.3349 \\ 2.8296 & -3.2975 \end{pmatrix}, \quad \tilde{D}_1 = \begin{pmatrix} 1.3029 & -0.1152 \\ 4.4041 & -0.3836 \end{pmatrix}.$$

The equality (2.3) is satisfied. However, $\tilde{\boldsymbol{\alpha}} \notin [0,1]$, $\widetilde{D_1}(1,2) = -0.1152 \ngeq 0$ and $\widetilde{D_1}(2,2) = -0.3836 \ngeq 0$, therefore the process defined by $\{\tilde{\boldsymbol{\alpha}}, \widetilde{D_0}, \widetilde{D_1}\}$ is not a real non-stationary MAP_2.

Example 2.2. *Consider the non-stationary MAP_2 given by*

$$\boldsymbol{\alpha} = (0.6865, 0.3135), \quad D_0 = \begin{pmatrix} -4.6962 & 4.2725 \\ 0.2059 & -0.3985 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.2746 & 0.1490 \\ 0.1091 & 0.0835 \end{pmatrix},$$

and the arbitrary invertible, stochastic matrix

$$B = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}.$$

From $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$, $\tilde{D}_0 = BD_0B^{-1}$ *and* $\tilde{D}_1 = BD_1B^{-1}$,

$$\tilde{\boldsymbol{\alpha}} = (0.9730, 0.0270), \quad \tilde{D}_0 = \begin{pmatrix} -6.3094 & 5.9550 \\ -1.4536 & 1.2148 \end{pmatrix}, \quad \tilde{D}_1 = \begin{pmatrix} 0.3082 & 0.0461 \\ 0.1888 & 0.0500 \end{pmatrix}.$$

The equality (2.3) is satisfied. However, $\widetilde{D}_0(2,1) = -1.4536 \ngeq 0$ *and* $\widetilde{D}_0(2,2) = 0.0500 \nleq 0$, *therefore the process defined by* $\{\tilde{\boldsymbol{\alpha}}, \widetilde{D}_0, \widetilde{D}_1\}$ *is not a real non-stationary* MAP_2.

Next we propose a procedure that shows how to build an infinite number of equivalent non-stationary MAP_2s for any given one. Different procedures to build infinite equivalent stationary MAPs for a given fixed one are provided in the literature, for instance Telek and Horváth [102] proceeded with an iterative numerical optimization method to find the matrix B for the similarity transform for any m-state MAP, and Ramírez-Cobo et al. [87] provided the conditions that two two-state stationary MAP representations need to satisfy in order to be equivalent. The following Theorem details how to select the elements of an invertible matrix $B = [\omega_1 \quad 1-\omega_1; \omega_2 \quad 1-\omega_2]$ such that the similarity transform presented in Theorem 2.1 always produces a well-defined equivalent non-stationary MAP_2.

Theorem 2.2. *Consider a non-stationary* MAP_2, $\mathcal{M}$, *as in (2.1), and define*

$$\begin{aligned}
\epsilon_1 &= \frac{(u-x+2z) - \sqrt{(u-x)^2 + 4zy}}{2(u-x+z-y)}, \\
\epsilon_2 &= \frac{(u+z+w-v) + \sqrt{(u+z+v+w)^2 - 4v(w+x+y)}}{2(u-x+z-y)}, \\
\epsilon_3 &= -\frac{v}{u+z}, \\
\epsilon_4 &= \frac{u-x-y}{(u-x-2y)}, \\
\epsilon_5 &= -\frac{v}{(x+y+w-v)}, \\
\epsilon_6 &= -\frac{v}{x+y}.
\end{aligned}$$

Without loss of generalization, assume that $x < u$ *and choose a value* ω_1 *from*

$$\max\{\alpha, \epsilon_1, \epsilon_2, \epsilon_3\} < \omega_1 < 1 \text{ if } z+u \neq 0 \text{ and } (u-x+z-y) \neq 0, \tag{2.5}$$

$$\max\{\alpha, \epsilon_1, \epsilon_2\} < \omega_1 < 1 \text{ if } z+u = 0 \text{ and } (u-x+z-y) \neq 0, \tag{2.6}$$

$$\max\{\alpha, \epsilon_4, \epsilon_5, \epsilon_6\} < \omega_1 < 1 \text{ if } (u-x+z-y) = 0. \tag{2.7}$$

Given ω_1, define

$$\begin{aligned}
\kappa_1 &= \frac{\omega_1(w-v)+v}{\omega_1(u-x+z-y)-z-u},\\
\kappa_2 &= \frac{\omega_1(z+u)-z}{\omega_1(u-x+z-y)-z+x},\\
\kappa_3 &= -\frac{\omega_1(w-v)+v}{x+y},\\
\kappa_4 &= \frac{\omega_1(x+y)-y+u-x}{u-y}.
\end{aligned}$$

Finally, choose ω_2 from

$$0 < \omega_2 < \min\{\alpha, \epsilon_1, \epsilon_2, \kappa_1, \kappa_2\} \text{ if } (u-x+z-y) \neq 0. \tag{2.8}$$

$$0 < \omega_2 < \min\{\alpha, \epsilon_4, \epsilon_5, \kappa_3, \kappa_4\} \text{ if } (u-x+z-y) = 0. \tag{2.9}$$

Then, there exists an infinite number of MAP_2s, $\widetilde{\mathcal{M}}$, given by $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}B^{-1}$, $\tilde{D}_0 = BD_0B^{-1}$ and $\tilde{D}_1 = BD_1B^{-1}$, where

$$B = \begin{pmatrix} \omega_1 & 1-\omega_1 \\ \omega_2 & 1-\omega_2 \end{pmatrix},$$

such that (2.3) holds.

See Appendix 2.A for the proof.

Remark 2.1. *We make some remarks regarding the denominator of ϵ_4, ϵ_5, ϵ_6 and κ_4 in the case where $(u-x+z-y)=0$, or equivalently, when $z=-u+x+y$.*

If $x+y=0$ then $\det(D_0)=0$, in contradiction with the invertibility of D_0. If $u-y=0$, then $z=x<0$, which is not a well-defined MAP_2, as it happens if $u-x-2y=0$, then $z=-y<0$. Finally, if $x+y+w-v=0$, then $v=x+y+w<0$, which again, is not a well-defined MAP_2.

We illustrate Theorem 2.2 by the following examples.

Example 2.3. *Consider the non-stationary MAP_2 given by*

$$\boldsymbol{\alpha} = (0.2537, 0.7463), \quad D_0 = \begin{pmatrix} -3.1255 & 0.4362 \\ 0.4877 & -0.9687 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.1284 & 1.5609 \\ 0.1526 & 0.3285 \end{pmatrix}.$$

where $u-x+z-y \neq 0$. By Theorem 2.2, a value ω_1 is chosen from

$$\max\{0.2537, 0.1781, 0.3978, 0.3171\} < \omega_1 < 1.$$

For example, consider $\omega_1 = 0.4699$*. Then, a second value* ω_2 *is chosen from*

$$0 < \omega_2 < \min\{0.2537, 0.1781, 0.3978, 0.4024, 0.2771\}.$$

For example, consider $\omega_2 = 0.0125$*. Then*

$$B = \begin{pmatrix} 0.4699 & 0.5301 \\ 0.0125 & 0.9875 \end{pmatrix}.$$

From $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$*,* $\tilde{D}_0 = BD_0B^{-1}$ *and* $\tilde{D}_1 = BD_1B^{-1}$*,*

$$\tilde{\boldsymbol{\alpha}} = (0.5274, 0.4726), \quad \tilde{D}_0 = \begin{pmatrix} -2.6042 & 1.0856 \\ 0.9812 & -1.4900 \end{pmatrix}, \quad \tilde{D}_1 = \begin{pmatrix} 1.2946 & 0.2241 \\ 0.3464 & 0.1624 \end{pmatrix}.$$

which is a well-defined non-stationary MAP_2 *equivalent to* $\{\boldsymbol{\alpha}, D_0, D_1\}$*.*

Example 2.4. *Consider the non-stationary* MAP_2 *given by*

$$\boldsymbol{\alpha} = (0.6984, 0.3016), \quad D_0 = \begin{pmatrix} -1.6255 & 1.5008 \\ 0.6235 & -0.7481 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0673 & 0.0573 \\ 0.0545 & 0.0702 \end{pmatrix}.$$

where $u - x + z - y = 0$*. By Theorem 2.2, a value* ω_1 *is chosen from*

$$\max\{0.6984, 0.2935, 0.4873, 0.4371\} < \omega_1 < 1.$$

For example, consider $\omega_1 = 0.699$*. Then, a second value* ω_2 *is chosen from*

$$0 < \omega_2 < \min\{0.6984, 0.2935, 0.4873, 0.5091, 0.3160\}.$$

For example, consider $\omega_2 = 0.2455$*. Then*

$$B = \begin{pmatrix} 0.6990 & 0.3010 \\ 0.2455 & 0.7545 \end{pmatrix}.$$

From $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$*,* $\tilde{D}_0 = BD_0B^{-1}$ *and* $\tilde{D}_1 = BD_1B^{-1}$*,*

$$\tilde{\boldsymbol{\alpha}} = (0.9986, 0.0014), \quad \tilde{D}_0 = \begin{pmatrix} -2.0243 & 1.8997 \\ 0.2247 & -0.3493 \end{pmatrix}, \quad \tilde{D}_1 = \begin{pmatrix} 0.0724 & 0.0522 \\ 0.0596 & 0.0650 \end{pmatrix}.$$

which is a well-defined non-stationary MAP_2 *equivalent to* $\{\boldsymbol{\alpha}, D_0, D_1\}$*.*

2.2 Canonical parametrization of the non-stationary MAP_2

The canonical representation of a process is of great importance when dealing with practical applications, for example, when the model is to be fitted to a real data. The existence of a unique representation of the problem avoids the typical switching problems and lack of convergence of the common statistical inference techniques, as the MLE, EM or Bayesian approaches. The previous Section showed that the usual representation of the non-stationary MAP_2 as in (2.1) is not unique. In the case of the stationary case, Bodrog et al. [7] found a unique, canonical representation that reduced the number of parameters from six to four. Such canonical representation depends on the value of a parameter γ, which is the eigenvalue different from 1 of $P^\star$, the probability transition matrix defined by (1.4). The parameters that defined the canonical version, $\{x_0, y_0, u_0, v_0\}$, are defined in terms of the first three moments of the stationary inter-event time distribution (T), μ_1, μ_2, μ_3 and the lag-one correlation coefficient ρ_1, given by

$$\begin{aligned} \mu_n &= E\left(T^n\right) = n!\boldsymbol{\phi}\left(-D_0\right)^{-n}\mathbf{e}, \quad \text{for } n = 1, 2, 3, \\ \rho_1 &= \gamma\frac{\frac{\mu_2}{2} - \mu_1^2}{\mu_2 - \mu_1^2}. \end{aligned} \tag{2.10}$$

The explicit expressions of $\{x_0, y_0, u_0, v_0\}$ can be found in Bodrog et al. [7] (page 469). They are obtained in terms of the three moments and the lag-one correlation coefficient, $\{\mu_1, \mu_2, \mu_3, \rho_1\}$, which characterize the stationary MAP_2.

In this Section, we derive the canonical representation of the non-stationary version of the MAP_2. Such representation is given by five parameters, instead of the seven values that characterized the MAP_2 as in (2.1). Note that in the stationary case, the initial probability $(\boldsymbol{\phi} = (\boldsymbol{\pi}D_1\mathbf{e})^{-1}\boldsymbol{\pi}D_1)$ is function of D_0 and D_1 while in the non-stationary case, the initial probability $(\boldsymbol{\alpha})$ is a new parameter independent of the rate matrices.

The next Theorem provides our main result.

Theorem 2.3. *The canonical representation for the non-stationary MAP_2 is unique and given by*

$$\boldsymbol{\alpha}^c = (\alpha_0,\ 1-\alpha_0), \quad D_0^c = \begin{pmatrix} x_0 & y_0 \\ 0 & u_0 \end{pmatrix}, \quad D_1^c = \begin{pmatrix} -x_0-y_0 & 0 \\ v_0 & -u_0-v_0 \end{pmatrix}, \tag{2.11}$$

if $\gamma > 0$. On the other hand, for those non-stationary MAP_2s such that $\gamma \leq 0$, then the *canonical form is*

$$\boldsymbol{\alpha}^c = (\alpha_0,\ 1-\alpha_0), \quad D_0^c = \begin{pmatrix} x_0 & y_0 \\ 0 & u_0 \end{pmatrix}, \quad D_1^c = \begin{pmatrix} 0 & -x_0 - y_0 \\ -u_0 - v_0 & v_0 \end{pmatrix}, \tag{2.12}$$

where $u_0 \leq x_0 < 0$, $y_0, v_0 \geq 0$, $x_0 + y_0 \leq 0$ *and* $u_0 + v_0 \leq 0$.

The initial probability α_0 *is defined as*

$$\alpha_0 = \frac{-x_0 - x_0 u_0 r_1}{u_0 - y_0 - x_0}, \tag{2.13}$$

where

$$r_1 = \frac{\alpha(-u - z + y + x) + z - x}{xu - yz}.$$

Proof. Since the non-stationary MAP_2 converges to its stationary version and its canonical form is unique, then from Bodrog et al. [7] we have that $\{D_0^c, D_1^c\}$ are well-defined rate matrices.

Given the structure of the matrices $\{D_0^c, D_1^c\}$ and the fact that any non-stationary MAP_2, $\mathcal{M}$, defined as in (2.1), is equivalent to its non-stationary canonical representation, then the equality of the moments of T_i and T_i^c holds,

$$\boldsymbol{\alpha}_i \left(-D_0\right)^{-m} \mathbf{e} = \boldsymbol{\alpha}_i^c \left(D_0^c\right)^{-m} \mathbf{e}, \text{ for all } i, m \geq 1, \tag{2.14}$$

where $\boldsymbol{\alpha}_i = \boldsymbol{\alpha} \left(P^\star\right)^{i-1}$ and $\boldsymbol{\alpha}_i^c = \boldsymbol{\alpha} \left(P^c\right)^{i-1}$ (being $P^c = \left(-D_0^c\right)^{-1} D_1^c$). Then, solving the system (2.14), we obtain that, for $z_0 = 0$, and for all $y_0 > 0$,

$$\mathcal{F}_0 = \begin{cases} u_0 = \frac{(u+x) - \sqrt{(u-x)^2 + 4yz}}{2}, \\ x_0 = \frac{(u+x) + \sqrt{(u-x)^2 + 4yz}}{2}, \\ \alpha_0 = \frac{-x_0 - x_0 u_0 r_1}{u_0 - y_0 - x_0}, \\ v_0 = \frac{x_0 u_0 (u_0(x - v - z + w) - (uw - xv - yv + zw + xu - yz))}{(xu - yz)(y_0 + x_0 - u_0)}, \\ w_0 = \frac{x_0((x_0 + y_0)(u_0(x - v - z + w) - xu + yz) - u_0(uw - xv - yv + zw))}{(xu - yz)(y_0 + x_0 - u_0)}, \end{cases} \tag{2.15}$$

where the set $\mathcal{F}_0$ solves the system of equations given by (2.1). We point out that the set $\mathcal{F}_0$ is analogous to the results provided by Bodrog et al. [7]. Note that $\det(D_0) = xu - yz \neq 0$.

The equations defined by $\mathcal{F}_0$ are used to define α_0, as stated in (2.13). It remains to prove that $0 \leq \alpha_0 \leq 1$. For the sake of brevity, it is given in Appendix 2.B.

We have proven that the canonical representation of the non-stationary MAP_2 (2.11) - (2.12) is characterized by $\{\alpha_0, x_0, y_0, u_0, v_0\}$ as in (2.15). We now proceed to prove the uniqueness of this representation. Assume we have two equivalent canonical non-stationary MAP_2,

$$\begin{aligned}\mathcal{M}_1 &= \{\boldsymbol{\alpha}_1^c, D_{0,1}^c, D_{1,1}^c\},\\ \mathcal{M}_2 &= \{\boldsymbol{\alpha}_2^c, D_{0,2}^c, D_{1,2}^c\},\end{aligned}$$

in either canonical form (2.11)-(2.12). Note that their corresponding canonical stationary versions, $\mathcal{M}_1^s = \{\boldsymbol{\phi}_1^c, D_{0,1}^c, D_{1,1}^c\}$ and $\mathcal{M}_2^s = \{\boldsymbol{\phi}_2^c, D_{0,1}^c, D_{1,2}^c\}$, are also equivalent. As the canonical representation for the stationary MAP_2 is unique, then $\mathcal{M}_1^s = \mathcal{M}_2^s$, that is,

$$\begin{aligned}\boldsymbol{\phi}_1^c &= \boldsymbol{\phi}_2^c,\\ D_{0,1}^c &= D_{0,2}^c,\\ D_{1,1}^c &= D_{1,2}^c.\end{aligned} \tag{2.16}$$

Let $\boldsymbol{\alpha}_1^c = (\alpha_{0,1}, 1-\alpha_{0,1})$. From (2.13), we can state that $\alpha_{0,1}$ is a function of r_1 and the parameter of $D_{0,1}^c$. Analogously with $\alpha_{0,2}$. From (2.16) and the fact that $\mathcal{M}_1$ and $\mathcal{M}_2$ are equivalent, we have that

$$\alpha_{0,1} = \alpha_{0,1}\{r_1; D_{0,1}^c\} = \alpha_{0,2}\{r_1; D_{0,2}^c\} = \alpha_{0,2}.$$

In consequence $\mathcal{M}_1 = \mathcal{M}_2 = \{\boldsymbol{\alpha}^c, D_0^c, D_1^c\}$, concluding that the canonical representation of the non-stationary MAP_2, given by Theorem 2.3, is unique.

□

Some examples of the canonical representation are as follows,

Example 2.5. *Consider the non-stationary MAP_2 given by*

$$\boldsymbol{\alpha} = (0.3022, 0.6978), \quad D_0 = \begin{pmatrix} -5.2715 & 0.4225 \\ 2.5818 & -6.5665 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.2858 & 3.5632 \\ 3.2469 & 0.7378 \end{pmatrix}.$$

It can be seen that $\gamma = -0.3168 < 0$, therefore the second canonical form (2.12) is computed as

$$\boldsymbol{\alpha}^c = (0.5718, 0.4282), \quad D_0^c = \begin{pmatrix} -4.6902 & 2.6175 \\ 0 & -7.1478 \end{pmatrix}, \quad D_1^c = \begin{pmatrix} 0 & 2.0726 \\ 5.1242 & 2.0236 \end{pmatrix}.$$

Example 2.6. *Consider the non-stationary MAP_2 given by*

$$\boldsymbol{\alpha} = (0.0334, 0.9666), \quad D_0 = \begin{pmatrix} -0.9491 & 0.1360 \\ 2.2415 & -3.4699 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.6856 & 0.1275 \\ 0.9939 & 0.2345 \end{pmatrix}.$$

In this case $\gamma = 0.0114 > 0$ and hence, the first canonical form (2.11) is computed,

$$\boldsymbol{\alpha}^c = (0.6825, 0.3175), \quad D_0^c = \begin{pmatrix} -0.8335 & 0.7948 \\ 0 & -3.5855 \end{pmatrix}, \quad D_1^c = \begin{pmatrix} 0.0386 & 0 \\ 2.7041 & 0.8815 \end{pmatrix}.$$

2.3 Characterization of the non-stationary $\boldsymbol{MAP_2}$

This section provides the starting point for the next chapter, which deals with the estimation of the parameters of the canonical non-stationary MAP_2 for a particular type of data set. Following Bodrog et al. [7], in this section we characterize the non-stationary MAP_2 in terms of a set of moments.

As stated before, for a given stationary MAP_2, Ramírez-Cobo et al. [87] provided an alternative method of characterizing the parameters of an equivalent stationary MAP_2 (without going through the similarity transform), following the conditions that two stationary MAP_2 need to have in order to be equivalent. They stated that if (2.3) is satisfied for $n = 1$ and $n = 2$, then it will hold for all $n \geq 3$. This result will help us to find the set of moments characterizing the non-stationary MAP_2.

Following Ramírez-Cobo et al. [87] approach, for the non-stationary MAP_2, we have that the following equalities for $n = 1$ and $n = 2$ must hold,

$$\begin{aligned} f^*_{T;\boldsymbol{\alpha},D_0,D_1}(s_1) &= f^*_{\tilde{T};\tilde{\boldsymbol{\alpha}},\tilde{D}_0,\tilde{D}_1}(s_1), \\ f^*_{T;\boldsymbol{\alpha},D_0,D_1}(s_1,s_2) &= f^*_{\tilde{T};\tilde{\boldsymbol{\alpha}},\tilde{D}_0,\tilde{D}_1}(s_1,s_2), \end{aligned}$$

or equivalently (from (1.7)),

$$\frac{s\epsilon + \nu}{s^2 + s\beta + \nu} = \frac{s\tilde{\epsilon} + \tilde{\nu}}{s^2 + s\tilde{\beta} + \tilde{\nu}}, \quad \text{for all } s, \tag{2.17}$$

and

$$\frac{s_1 s_2 \delta_1 + s_2 \delta_2 + s_1 \epsilon \nu + \nu^2}{s_1^1 s_2^1 + (s_1^2 + s_2^2)\nu + (s_1^2 s_2 + s_1 s_2^2)\beta + (s_1 + s_2)\beta\nu + s_1 s_2 \beta^2 + \nu^2} = \tag{2.18}$$
$$\frac{s_1 s_2 \tilde{\delta}_1 + s_2 \tilde{\delta}_2 + s_1 \tilde{\epsilon}\tilde{\nu} + \tilde{\nu}^2}{s_1^1 s_2^1 + (s_1^2 + s_2^2)\tilde{\nu} + (s_1^2 s_2 + s_1 s_2^2)\tilde{\beta} + (s_1 + s_2)\tilde{\beta}\tilde{\nu} + s_1 s_2 \tilde{\beta}^2 + \tilde{\nu}^2}, \quad \text{for all } s_1 \text{ and } s_2,$$

where ϵ, β, ν, δ_1, δ_2 (respectively $\widetilde{\epsilon}$, $\widetilde{\beta}$, $\widetilde{\nu}$, $\widetilde{\delta}_1$, $\widetilde{\delta}_2$) are given by

$$\begin{aligned}
\epsilon &= \alpha(z+u-x-y)-(z+u), \\
\beta &= -x-u, \\
\nu &= xu-yz, \\
\delta_1 &= \alpha(z+u-x-y)(-z-u+w-v)+(z+u-x-y)v+(z+u)^2, \\
\delta_2 &= \alpha(x+y-z-u)(uw-yv-xv+zw)+(x+y-z-u)(xv-zw)-(u+z)\nu.
\end{aligned} \tag{2.19}$$

Ramírez-Cobo and Lillo [91] state that the equality (2.17) has one solution if the value τ, defined by

$$\tau = \nu + \epsilon(\epsilon - \beta),$$

is found to be different from zero. Otherwise, there will exist an infinite number of solutions (see page 6 of [91] for the explicit expressions). Therefore, if $\tau \neq 0$, we have that the equalities of the Laplace transforms for $n = 1$ and $n = 2$, given by (2.17) and (2.18) respectively, will hold if and only if

$$\tilde{\epsilon} = \epsilon, \quad \tilde{\beta} = \beta, \quad \tilde{\nu} = \nu, \quad \tilde{\delta}_1 = \delta_1, \quad \tilde{\delta}_2 = \delta_2. \tag{2.20}$$

We have introduced all the results needed to prove the next result. The following Proposition provides the set of five moments that fully characterized the canonical non-stationary MAP_2. But first, recall that the moments of T_i are defined in (1.6), where

$$\mu_{i,m} = E\left(T_i^m\right) = m!\boldsymbol{\alpha}_i\left(-D_0\right)^{-m}\mathbf{e},$$

with $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}\left(P^\star\right)^{i-1}$.

Proposition 2.1. *Consider two non-stationary MAP_2s, $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ defined as in (2.1), such that $\tau \neq 0$. The equality of the Laplace transforms (2.3) holds if and only if the following equalities hold*

$$\begin{aligned}
r_1(\mathcal{M}) &= r_1(\widetilde{\mathcal{M}}), \\
r_2(\mathcal{M}) &= r_2(\widetilde{\mathcal{M}}), \\
r_3(\mathcal{M}) &= r_3(\widetilde{\mathcal{M}}), \\
\mu_2(\mathcal{M}) &= \mu_2(\widetilde{\mathcal{M}}), \\
\mu_3(\mathcal{M}) &= \mu_3(\widetilde{\mathcal{M}}),
\end{aligned} \tag{2.21}$$

where

$$
\begin{aligned}
r_1(\mathcal{M}) &= \mu_{1,1} = \boldsymbol{\alpha}\,(-D_0)^{-1}\,\mathbf{e}, \\
r_2(\mathcal{M}) &= \frac{\mu_{1,2}}{2} = \boldsymbol{\alpha}\,(-D_0)^{-2}\,\mathbf{e}, \\
r_3(\mathcal{M}) &= \frac{\mu_{1,3}}{3!} = \boldsymbol{\alpha}\,(-D_0)^{-3}\,\mathbf{e}, \\
\mu_2(\mathcal{M}) &= \mu_{2,1} = \boldsymbol{\alpha}\,(P^\star)\,(-D_0)^{-1}\,\mathbf{e}, \\
\mu_3(\mathcal{M}) &= \mu_{3,1} = \boldsymbol{\alpha}\,(P^\star)^2\,(-D_0)^{-1}\,\mathbf{e},
\end{aligned} \tag{2.22}
$$

$r_1(\widetilde{\mathcal{M}})$, $r_2(\widetilde{\mathcal{M}})$ $r_3(\widetilde{\mathcal{M}})$, $\mu_2(\widetilde{\mathcal{M}})$, $\mu_3(\widetilde{\mathcal{M}})$ *are defined analogously.*

Remark 2.2. *Notice that in the non-stationary case, the inter-event time distribution, T_i, has a differently parametrize PH-distribution for each $i \geq 1$, therefore we have no auto-correlation function (and no $\rho(1)$), as in the case of the stationary MAP$_2$, but instead the correlations between T_is.*

Proof. Assume that the equality of the Laplace transforms (2.3) holds for $n = 1$ and $n = 2$. The expressions in (2.22) can be written in terms of (2.19) as,

$$
\begin{aligned}
r_1(\mathcal{M}) &= \frac{\beta - \epsilon}{\nu}, \\
r_2(\mathcal{M}) &= \frac{\beta r_1 - 1}{\nu}, \\
r_3(\mathcal{M}) &= \frac{\beta r_2 - r_1}{\nu}, \\
\mu_2(\mathcal{M}) &= \frac{\beta\nu - \delta_2}{\nu^2}, \\
\mu_3(\mathcal{M}) &= \frac{\delta_1(\beta - \epsilon - \mu_2\nu) + \mu_2\nu(\epsilon\beta - \mu_2\epsilon\nu + \nu) + \epsilon\beta(\epsilon - \beta)}{\nu\tau}.
\end{aligned} \tag{2.23}
$$

Therefore it can easily be checked that if (2.20) holds, then the equalities (2.21) will also hold.

Consider now that the equalities given by (2.21) are satisfied. From the duality established in (2.23),

$$\begin{aligned}\epsilon &= \beta - r_1\nu, &(2.24)\\ \nu &= \frac{r_1\beta - 1}{r_2}, &(2.25)\\ \nu &= \frac{r_2\beta - r_1}{r_3}, &(2.26)\\ \delta_2 &= \beta\nu - \mu_2\nu^2, &(2.27)\\ \delta_1 &= \frac{\mu_3\nu(\nu+\epsilon^2-\epsilon\beta) - \mu_2\nu(\epsilon\beta - \mu_2\epsilon\nu + \nu) - \epsilon\beta(\epsilon-\beta)}{(\beta-\epsilon-\mu_2\nu)}. &(2.28)\end{aligned}$$

Then, from (2.21) and the equality between (2.25) and (2.26), we obtain that

$$\beta = \frac{r_3(\mathcal{M}) - r_1(\mathcal{M})r_2(\mathcal{M})}{r_1(\mathcal{M})r_3(\mathcal{M}) - r_2(\mathcal{M})^2} = \frac{r_3(\widetilde{\mathcal{M}}) - r_1(\widetilde{\mathcal{M}})r_2(\widetilde{\mathcal{M}})}{r_1(\widetilde{\mathcal{M}})r_3(\widetilde{\mathcal{M}}) - r_2(\widetilde{\mathcal{M}})^2} = \widetilde{\beta} \qquad (2.29)$$

From (2.25) and (2.24) we obtain that $\nu = \widetilde{\nu}$ and $\epsilon = \widetilde{\epsilon}$. This, in combination with (2.27) and (2.28) yields $\delta_2 = \widetilde{\delta_2}$ and $\delta_1 = \widetilde{\delta_1}$. Since (2.20) is satisfied, the equality of the Laplace transforms (2.3) will hold for $n \geq 1$. □

Remark 2.3. *In order to show that the set equations (2.23) and equations (2.28) and (2.29) are well-defined, note that D_0 is non-singular, which implies that $\nu \neq 0$. Besides, since $\tau \neq 0$, then $\nu\tau \neq 0$, which are the denominators of the set of expressions given in (2.23). Consider now the denominator of (2.28), where*

$$\begin{aligned}\beta - \epsilon - \mu_2\nu &= \nu(r_1 - \mu_2)\\ &= \frac{(y-u-z+x)\left[\alpha(-xv+xu+uw-yv-yz+zw) - zw + xv\right]}{\nu},\end{aligned}$$

since

$$\tau = (y-u-z+x)\Big[(y-u-z+x)\alpha^2 + (-x+u+2z)\alpha - z\Big] \neq 0,$$

which implies that $(y-u-z+x) \neq 0$. Also, $[\alpha(-xv+xu+uw-yv-yz+zw) - zw + xv] \neq 0$ is equivalent to

$$\alpha \neq \frac{zw - xv}{-xv+xu+uw-yv-yz+zw},$$

which is equivalent to $\boldsymbol{\alpha} \neq \boldsymbol{\phi}$, and this holds since in this work we focus on the non-stationary version of the MAP_2. Therefore $\beta - \epsilon - \mu_2\nu \neq 0$ always holds.

Finally, we prove that $r_1r_3 - r_2^2 \neq 0$, which is the denominator of (2.29), where

$$r_1r_3 - r_2^2 = -\frac{\tau}{\nu^3} \neq 0.$$

Solving the system of equations given by (2.21), we have an alternative way of obtaining the parameters of an equivalent non-stationary MAP_2.

2.4 Chapter summary

We have examined the identifiability of the non-stationary MAP_2 process in detail. The identifiability of the process is a crucial property, since it determines the possible multimodality of the likelihood function. This fact has to be taken into consideration in order to develop an estimation method to fit the model to real data sets.

In this chapter we prove that the m-state non-stationary MAP is non-identifiable. We present a procedure that shows how to construct an equivalent non-stationary MAP_2s to any given fixed non-stationary MAP_2, which is based on the definition of a similarity transform matrix B. We also show that the non-stationary MAP_2 is characterized by a set of five moments. The main result of this chapter is the proof of the existence of a unique, canonical representation for the non-stationary MAP_2, which is a powerful result that eases the estimation of the process parameters. We have illustrated our findings with numerical examples as well.

Unlike the stationary version of the process, in the non-stationary MAP the inter-event times are not identically distributed. This fact makes the non-stationary MAP to have more applicability in terms of modeling than its stationary counterpart, as we will see in the following Chapter.

Appendix

2.A Proof of Theorem 2.2

The similarity transform, $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} B^{-1}$, $\tilde{D}_0 = BD_0B^{-1}$ and $\tilde{D}_1 = BD_1B^{-1}$, where B is given as in Theorem 2.2 provides the following parameter values

$$\begin{aligned}
\tilde{\alpha} &\equiv \frac{\alpha-\omega_2}{\omega_1-\omega_2},\\
\tilde{y} &\equiv \frac{-\omega_1^2(u-x+z-y)+\omega_1(u+2z-x)-z}{\omega_1-\omega_2},\\
\tilde{z} &\equiv \frac{(\omega_2^2(u-x+z-y)-\omega_2(u+2z-x)+z)}{\omega_1-\omega_2},\\
\tilde{v} &\equiv \frac{(-\omega_2^2(u-x+z-y)+\omega_2(u+z+w-v)+v)}{\omega_1-\omega_2},\\
\tilde{x} &\equiv \frac{(\omega_1\omega_2(u-x+z-y)+\omega_1(x-z)-\omega_2(z+u)+z)}{\omega_1-\omega_2},\\
\tilde{u} &\equiv \frac{(-\omega_1\omega_2(u-x+z-y)-\omega_2(x-z)+\omega_1(z+u)-z)}{\omega_1-\omega_2},\\
\tilde{w} &\equiv \frac{(-\omega_1\omega_2(u-x+z-y)+\omega_1(w-v)+\omega_2(z+u)+v)}{\omega_1-\omega_2},\\
\tilde{s}_1 &\equiv \frac{(\omega_1^2(u-x+z-y)-\omega_1(u+z+w-v)-v)}{\omega_1-\omega_2},\\
\tilde{s}_2 &\equiv \frac{-(-\omega_1\omega_2(u-x+z-y)+\omega_2(w-v)+\omega_1(z+u)+v)}{\omega_1-\omega_2},
\end{aligned} \tag{2.30}$$

where

$$\begin{aligned}
\tilde{s}_1 &= (-\tilde{x}-\tilde{y}-\tilde{w}),\\
\tilde{s}_2 &= (-\tilde{z}-\tilde{u}-\tilde{v}).
\end{aligned}$$

We start by proving that the set (2.30) is well-defined. Let us first assume that $x<u$. The proof will be divided into two parts, one per each case $(u-x+z-y)\neq 0$ and $(u-x+z-y)=0$.

If $\omega_2<\alpha<\omega_1$, then $\omega_1-\omega_2>0$ and $\tilde{\alpha}\in[0,1]$ will hold for both cases. Consider first the case $(u-x+z-y)\neq 0$.

Since,

$$\omega_2 < \frac{(u-x+2z)-\sqrt{(u-x)^2+4zy}}{2(u-x+z-y)} < \omega_1$$

it implies that $\tilde{y}>0$ and $\tilde{z}>0$.

In addition,

$$\omega_2 < \frac{(u+z+w-v)+\sqrt{(u+z+v+w)^2-4v(w+x+y)}}{2(u-x+z-y)} < \omega_1,$$

which implies that $\tilde{s}_1>0$ and $\tilde{v}>0$.

Moreover,

$$\omega_2 < \frac{\omega_1(z+u)-z}{\omega_1(u-x+z-y)-z+x} < -\frac{\omega_1(x-z)+z}{\omega_1(u-x+z-y)-z-u},$$

which implies that $\tilde{x} < 0$ and $\tilde{u} < 0$.

We now prove that $\tilde{w} > 0$ and $\tilde{s}_2 > 0$. By imposing the condition that

$$-\frac{v}{u+z} < \omega_1, \tag{2.31}$$

when $u+z \neq 0$, we obtain the following. If $\omega_1(u-x+z-y)-(w-v) > 0$, then

$$\frac{\omega_1(u+z)+v}{\omega_1(u-x+z-y)-(w-v)} < \omega_2 < \frac{\omega_1(w-v)+v}{\omega_1(u-x+z-y)-z-u},$$

which implies $\tilde{s}_2 > 0$ and $\tilde{w} > 0$. Now, if $\omega_1(u-x+z-y)-(w-v) < 0$,

$$\omega_2 < \frac{\omega_1(w-v)+v}{\omega_1(u-x+z-y)-z-u} < \frac{\omega_1(u+z)+v}{\omega_1(u-x+z-y)-(w-v)},$$

assures that $\tilde{s}_2 > 0$ and $\tilde{w} > 0$ are fulfilled. If $z+u=0$ then it is not necessary to impose condition (2.31).

It remains to prove that $\max\{\epsilon_2, \epsilon_3, \epsilon_4\} < 1$ and $0 < \min\{\epsilon_1, \epsilon_2, \kappa_1, \kappa_2\}$.

If $(u-x+z-y) > 0$, then

$$\begin{aligned}
(2(u-x+z-y)-(u-x+2z))^2 &= (u-x)^2+4y(y-u+x) \\
&< (u-x)^2+4yz, \\
(u-x+2z)^2 &= (u-x)^2+4z(z+u-x) \\
&> (u-x)^2+4zy,
\end{aligned}$$

and

$$\begin{aligned}
(2(u-x+z-y)-(u+z+w-v))^2 &= (u+z+v+w)^2-4v(x+y+w) \\
&\quad -4(u-x+z-y)(x+y+w) \\
&> (u+z+v+w)^2-4v(x+y+w), \\
(u+z+w-v)^2 &= (u+z+v+w)^2-4v(x+y+w)-4v(u-x+z-y) \\
&< (u+z+v+w)^2-4v(x+y+w),
\end{aligned}$$

which proves that $0 < \epsilon_1 < 1$ and $0 < \epsilon_2 < 1$. Now, if $(u-x+z-y) < 0$,

$$\begin{aligned}
(2(u-x+z-y)-(u-x+2z))^2 &= (u-x)^2 + 4y(y-u+x) \\
&> (u-x)^2 + 4yz, \\
(u+z+w-v)^2 &= (u+z+v+w)^2 - 4v(x+y+w) - 4v(u-x+z-y) \\
&> (u+z+v+w)^2 - 4v(x+y+w),
\end{aligned}$$

and

$$\begin{aligned}
(2(u-x+z-y)-(u+z+w-v))^2 &= (u+z+v+w)^2 - 4v(x+y+w) \\
&\quad -4(u-x+z-y)(x+y+w) \\
&< (u+z+v+w)^2 - 4v(x+y+w), \\
(u-x+2z)^2 &= (u-x)^2 + 4z(z+u-x) \\
&< (u-x)^2 + 4zy,
\end{aligned}$$

then $0 < \epsilon_1 < 1$ and $0 < \epsilon_2 < 1$ are proven. In addition, $\epsilon_3 < 1$ since $-z-u-v > 0$. Therefore, we obtain that $\{\epsilon_1, \epsilon_2, \epsilon_3\} < 1$ and $\{\epsilon_1, \epsilon_2\} > 0$.

Recall that

$$\begin{aligned}
\kappa_1 &= \frac{\omega_1(w-v)+v}{\omega_1(u-x+z-y)-z-u}, \\
\kappa_2 &= \frac{\omega_1(z+u)-z}{\omega_1(u-x+z-y)-z+x}.
\end{aligned}$$

If $w-v \geq 0$, then $\omega_1(w-v)+v > 0$. If $w-v<0$ then $\omega_1(w-v)+v>0$ since $\omega_1 < 1$. In addition, if $(u-x+z-y) > 0$ then $\omega_1(u-x+z-y)-z-u > 0$, since $-z-u > 0$. If $(u-x+z-y) < 0$,

$$\omega_1(u-x+z-y)-z-u > (u-x+z-y)-z-u = -x-y > 0.$$

Therefore, $\kappa_1 > 0$. Finally, if $(u-x+z-y) > 0$, then $\omega_1(u-x+z-y)-z+x < (u-y) < 0$. If $(u-x+z-y) < 0$,

$$\omega_1(u-x+z-y)-z+x < \epsilon_1(u-x+z-y)-z+x = \frac{(u+x)-\sqrt{(u-x)^2+4zy}}{2} < 0,$$

and the fact that $z+u<0$ and $\omega_1 > \epsilon_1 > 0$, we obtain that $\omega_1(z+u)-z<0$, therefore $\kappa_2 > 0$.

Continue with the case $(u - x + z - y) = 0$, or equivalently $z = -u + x + y$. Then (2.30) becomes

$$\begin{aligned}
\tilde{y} &\equiv \frac{\omega_1(x - u + 2y) + u - x - y}{\omega_1 - \omega_2}, \\
\tilde{z} &\equiv \frac{(-\omega_2(x - u + 2y) - u + x + y)}{\omega_1 - \omega_2}, \\
\tilde{v} &\equiv \frac{(\omega_2(x + y + w - v) + v)}{\omega_1 - \omega_2}, \\
\tilde{x} &\equiv \frac{(\omega_1(u - y) - \omega_2(x + y) - u + x + y)}{\omega_1 - \omega_2}, \\
\tilde{u} &\equiv \frac{(-\omega_2(u - y) + \omega_1(x + y) + u - x - y)}{\omega_1 - \omega_2}, \\
\tilde{w} &\equiv \frac{(\omega_1(w - v) + \omega_2(x + y) + v)}{\omega_1 - \omega_2}, \\
\tilde{s}_1 &\equiv \frac{(-\omega_1(x + y + w - v) - v)}{\omega_1 - \omega_2}, \\
\tilde{s}_2 &\equiv \frac{-(\omega_2(w - v) + \omega_1(x + y) + v)}{\omega_1 - \omega_2}.
\end{aligned} \tag{2.32}$$

Now,

$$\omega_2 < \frac{u - x - y}{(u - x - 2y)} < \omega_1$$

implies that $\tilde{y} > 0$ and $\tilde{z} > 0$.

In addition,

$$\omega_2 < -\frac{v}{(x + y + w - v)} < \omega_1$$

implies that $\tilde{s}_1 > 0$ and $\tilde{v} > 0$.

Moreover,

$$\omega_2 < \frac{\omega_1(x + y) - y + u - x}{u - y} < \frac{\omega_1(u - y) - u + x + y}{x + y}$$

implies that $\tilde{x} < 0$ and $\tilde{u} < 0$.

We now prove that $\tilde{w} > 0$ and $\tilde{s}_2 > 0$. By imposing the condition that

$$-\frac{v}{x + y} < \omega_1,$$

we obtain the following. If $(w - v) \geq 0$, then

$$\omega_2 < \frac{\omega_1(w - v) + v}{x + y} < \frac{\omega_1(x + y) + v}{w - v},$$

which implies $\tilde{s}_2 > 0$ and $\tilde{w} > 0$. Now, if $(w - v) < 0$,

$$\frac{\omega_1(x+y)+v}{w-v} < \omega_2 < \frac{\omega_1(w-v)+v}{x+y},$$

which assures that $\tilde{s}_2 > 0$ and $\tilde{w} > 0$ are fulfilled.

It remains to prove that $\max\{\epsilon_4, \epsilon_5, \epsilon_6\} < 1$ and $0 < \min\{\epsilon_4, \epsilon_5, \kappa_3, \kappa_4\}$.

Since $z = -u + x + y > 0$, we have that $u - x - y < 0$ and $u - x - 2y < 0$. Moreover, $u-x-y > u-x-2y$, which implies $0 < \epsilon_4 < 1$. Since $-x-y-w > 0$, then $x+y+w-v < 0$. In addition, $-v > x + y + w - v0$, which implies $0 < \epsilon_5 < 1$. Furthermore, $\epsilon_6 < 1$ since $-z - u - v = x + y - v > 0$.

Recall that

$$\begin{aligned} \kappa_3 &= -\frac{\omega_1(w-v)+v}{x+y}, \\ \kappa_4 &= \frac{\omega_1(x+y)-y+u-x}{u-y}. \end{aligned}$$

If $w - v \leq 0$, then $\omega_1(w - v) + v > 0$, $\omega_1 < 1$. If $w - v > 0$ then $\omega_1(w - v) + v > 0$ since $\omega_1 > \epsilon_6$. Therefore $\kappa_3 > 0$, since $x + y < 0$. Finally,

$$\omega_1(x+y) - y + u - x < \epsilon_4(x+y) - y + u - x = \frac{(u-x-y)(u-y)}{u-x-2y} > 0,$$

since $u - y < 0$. Therefore $\kappa_4 > 0$, which ends the proof.

2.B Proof of $0 \leq \alpha_0 \leq 1$

The value of α_0 is obtained from (2.13) as

$$\alpha_0 = \frac{-x_0 - x_0 u_0 r_1}{u_0 - y_0 - x_0},$$

where

$$r_1 = \frac{\alpha(-u-z+y+x)+z-x}{xu-yz} > 0. \tag{2.33}$$

In order to show that $0 \leq \alpha_0$, it is sufficient to prove that $-x_0 - r_1(x_0 u_0) < 0$ and $u_0 - y_0 - x_0 < 0$ or equivalently $x_0 + r_1(x_0 u_0) > 0$ and $-u_0 + y_0 + x_0 > 0$.

Consider the first expression $x_0 + r_1(x_0 u_0) > 0$. Since $x_0 < 0$, then

$$x_0 + r_1(x_0 u_0) > 0 \Leftrightarrow (1 + r_1 u_0) < 0 \Leftrightarrow r_1 > -\frac{1}{u_0} \quad (u_0 < 0).$$

So, we have to prove that $r_1 > -\frac{1}{u_0}$. From (2.15),

$$\begin{aligned} -u_0 &= \frac{-(u+x)+\sqrt{(u-x)^2+4yz}}{2} \\ &> \frac{-(u+x)+\sqrt{(u-x)^2}}{2} \\ &= \frac{-(u+x)-(u-x)}{2} = -u. \end{aligned}$$

Then

$$0 < -u < -u_0. \tag{2.34}$$

From (2.33),

$$r_1 > \frac{-x}{xu-yz}. \tag{2.35}$$

Combining (2.34) and (2.35),

$$r_1(-u_0) > r_1(-u) > \frac{xu}{xu-yz} > 1. \tag{2.36}$$

Therefore, $r_1(-u_0) > 1$ if and only of $r_1 > -\frac{1}{u_0}$.

Second, consider $y_0+x_0-u_0 > 0$. We have that $x_0-u_0 > 0$ and $y_0 > 0$, then $y_0+x_0-u_0 > 0$, which implies that $0 \leq \alpha_0$.

We now proceed to prove that $\alpha_0 \leq 1$, which is equivalent to

$$r_1(x_0u_0) + u_0 - y_0 \leq 0. \tag{2.37}$$

Note that from (2.15),

$$\begin{aligned} -u_0 &= \frac{-(u+x)+\sqrt{(u-x)^2+4yz}}{2}, \\ -x_0 &= \frac{-(u+x)-\sqrt{(u-x)^2+4yz}}{2}, \end{aligned}$$

therefore

$$xu - yz = x_0u_0, \tag{2.38}$$
$$u + x = u_0 + x_0. \tag{2.39}$$

From Bodrog et al. [7] (see Definition 1),

$$x_0 + y_0 = x_0 - (1-a)x_0 = ax_0, \tag{2.40}$$

where $0 \leq a \leq 1$.

Then, by substituting (2.38, (2.39) and (2.40) in (2.37), we obtain

$$\begin{aligned} r_1(x_0u_0) + u_0 - y_0 &= \alpha(-z-u+x+y) + z - x + u + x - x_0 - y_0 \\ &= \alpha(-z-u+x+y) + z + u - x_0 - y_0 \\ &= -\epsilon - ax_0, \end{aligned}$$

where

$$\epsilon = \alpha(z+u-x-y) - (z+u).$$

Therefore, (2.37) has been simplified to

$$-\epsilon - ax_0 \leq 0. \tag{2.41}$$

Since $\epsilon = \alpha(z+u-x-y) - (z+u)$, we divide the proof into two cases: $(z+u-x-y) \leq 0$ and $(z+u-x-y) > 0$.

Assume first that $(z+u-x-y) < 0$ is satisfied. Then,

$$-x-y \leq \epsilon \leq -z-u. \tag{2.42}$$

Hence, (2.41) and (2.42) lead to

$$-\epsilon - ax_0 \leq x + y - ax_0 \leq 0.$$

Therefore, in this case, we have to prove that

$$x + y - ax_0 \leq 0. \tag{2.43}$$

Since the parameter a depends on the sign of γ, which is the eigenvalue different from 1 of $P^\star$ (recall that $P^\star = (-D_0)^{-1}D_1$). We consider both cases: $\gamma \leq 0$ and $\gamma > 0$. We begin with $\gamma \leq 0$. Then,

$$a = \frac{-\gamma}{p(1-\gamma) - \varphi\gamma}, \tag{2.44}$$

where

$$\begin{aligned} p &= \frac{-x_0(xu - yz + uw - yv - xv + zw) + (xu - yz)(-z + x + w - v)}{u_0(xu - yz + uw - yv - xv + zw)}, \\ \gamma &= \frac{-(uw - yv - xv + zw)}{xu - yz}, \\ \varphi &= \frac{x_0}{u_0}. \end{aligned} \tag{2.45}$$

The parameter p is analogously defined in Bodrog et al. [7], but in this case it has been rewritten according to our parametrization (2.1) and (2.11)-(2.12) (both definitions of p are numerically consistent).

After some straightforward calculations,

$$ax_0 = \frac{-\gamma x_0}{p(1-\gamma)-\varphi\gamma} = \frac{-(uw-yv-xv+zw)}{(-z+x+w-v-x_0)}. \tag{2.46}$$

Now

$$\begin{aligned} ax_0-(x+y) &= \frac{-(uw-yv-xv+zw)}{(-z+x+w-v-x_0)}-(x+y) \\ &= \frac{-(uw-yv-xv+zw)-(x+y)(-z+x+w-v-x_0)}{(-z+x+w-v-x_0)} \\ &= \frac{w(u+z)-(x+y)(-z+x+w-x_0)}{(-z+x+w-v-x_0)}. \end{aligned}$$

Taking into account that $(z+u-x-y) \leq 0$, $x-x_0<0$ and $x+y<0$,

$$\begin{aligned} w(u+z)-(x+y)(-z+x+w-x_0) &\leq w(x+y)-(x+y)(-z+x+w-x_0) \\ &= -(x+y)(-z+x-x_0) \\ &< -(x+y)(-z) \leq 0. \end{aligned}$$

On the other hand, since $ax_0<0$ and $\gamma<0$, then $(-z+x+w-v-x_0)<0$. Therefore we conclude that (2.43) is satisfied.

We consider now the case $\gamma>0$. Then,

$$a = \frac{1}{2\varphi}\left(1+\varphi\gamma-p(1-\gamma)-\sqrt{(1+\varphi\gamma-p(1-\gamma))^2-4\varphi\gamma}\right), \tag{2.47}$$

where p, γ and φ are defined in (2.45).

As before, we performed the corresponding calculations, and obtained

$$ax_0 = \frac{1}{2}\left((u+z+v-w)+\sqrt{(u+z+v-w)^2-4(uw-yv-xv+zw)}\right), \tag{2.48}$$

which implies that,

$$ax_0-(x+y) = \frac{1}{2}\left((u+z+v-w)-2(x+y)+\sqrt{(u+z+v-w)^2-4(uw-yv-xv+zw)}\right).$$

Note that

$$(u+z+v-w)-2(x+y)=(u+z+v+w)-2(x+y+w), \tag{2.49}$$

$$(u+z+v-w)^2-4(uw-yv-xv+zw)=(u+z+v+w)^2-4v(x+y+w). \tag{2.50}$$

If $(u+z+v-w)-2(x+y)\geq 0$, then (2.43) is fulfilled, since (2.50) is always non-negative. Now, if $(u+z+v-w)-2(x+y)<0$, then from (2.50) and the fact that (2.49) implies that $(u+z+v-w)<0$, we obtain that

$$\begin{aligned}
&(u+z+v-w)-2(x+y)+\sqrt{(u+z+v-w)^2-4(uw-yv-xv+zw)}\\
&\quad\geq (u+z+v+w)-2(x+y+w)+\sqrt{(u+z+v+w)^2}\\
&\quad= (u+z+v+w)-2(x+y+w)+|(u+z+v+w)|\\
&\quad= (u+z+v+w)-2(x+y+w)-(u+z+v+w)\\
&\quad= -2(x+y+w)\geq 0,
\end{aligned}$$

concluding that (2.43) is also fulfilled.

Consider now the case $(z+u-x-y)>0$. Then, similarly as in (2.42)

$$-x-y\geq\epsilon\geq -z-u. \tag{2.51}$$

Hence, from (2.41) and (2.51),

$$-\epsilon-ax_0\leq z+u-ax_0\leq 0.$$

Therefore, in this case, we need to prove that

$$z+u-ax_0\leq 0. \tag{2.52}$$

As before, we will consider the cases $\gamma\leq 0$ and $\gamma>0$. Assume first that $\gamma\leq 0$ and consider a and x_0 as defined in (2.44) and (2.46), respectively. Then

$$\begin{aligned}
ax_0-(z+u) &= \frac{-(uw-yv-xv+zw)}{(-z+x+w-v-x_0)}-(z+u),\\
&= \frac{-v(x+y)-(z+u)(-z+x-v-x_0)}{(-z+x+w-v-x_0)}.
\end{aligned}$$

Taking into account that $(z+u-x-y)\geq 0$, $x-x_0<0$ and $x+y<0$, we have that

$$\begin{aligned}
-v(x+y)-(z+u)(-z+x-v-x_0) &< -v(x+y)-(x+y)(-z+x-v-x_0)\\
&= -(x+y)(-z+x-x_0)\\
&< -(x+y)(-z)\leq 0.
\end{aligned}$$

Since it was already checked that $(-z+x+w-v-x_0)<0$, we can conclude that (2.52) is satisfied.

Assume now that $\gamma > 0$. As in the previous case, a and ax_0 were defined in (2.47) and (2.48), respectively.Then

$$ax_0-(z+u) = \frac{1}{2}\left((u+z+v-w)-2(z+u)+\sqrt{(u+z+v-w)^2-4(uw-yv-xv+zw)}\right).$$

In this case, note that

$$(u+z+v-w)-2(z+u) = (v-w-z-u), \tag{2.53}$$
$$(u+z+v-w)^2-4(uw-yv-xv+zw) = (v-w-z-u)^2+4v(u+z-y-x). \tag{2.54}$$

If $(u+z+v-w)-2(z+u)\geq 0$, then (2.52) is fulfilled, since (2.54) is always non-negative. Now, if $(u+z+v-w)-2(z+u)<0$, then from (2.54) and the fact that (2.53) implies that $(v-w-z-u)<0$, we obtain that

$$\begin{aligned}
&(u+z+v-w)-2(x+y)+\sqrt{(u+z+v-w)^2-4(uw-yv-xv+zw)}\\
&\quad\geq (v-w-z-u)+\sqrt{(v-w-z-u)^2},\\
&\quad= (v-w-z-u)+|(u+z+v+w)|,\\
&\quad= (v-w-z-u)-(u+z+v+w)=0,
\end{aligned}$$

concluding that (2.52) is also fulfilled.

We have that (2.41) is satisfied when $(z+u-x-y)\leq 0$ and $(z+u-x-y)>0$, which implies that $\alpha_0\leq 1$ and the proof is completed.

Chapter 3

Failure modeling of an electrical N-component framework by the non-stationary MAP_2

Electrical components are essential in everyday operations and life and it is crucial that they do not fail, even though they can break or malfunction at any time. Reliability in this context is described as the probability of a system or a component to function under stated conditions for a specified period of time [45]. Failures can be caused by faults or errors in the components that comprise the system, or alternatively, the structure that comprises the component. They can be due to internal (wearing out of the mechanism) or external factors (voltage, strength, vibrations, environmental conditions). When the lifetime and deterioration of a system or a component is submitted to external failures, it is usually referred as shock models, while we refer to wear models when the failures are due to internal factors. As a failure occurs, a repair or replacement may take place in order that the component goes back to functioning as soon as possible.

Reliability and maintenance policies of systems considering random component failure have been widely studied in the literature, see [24, 31, 57, 68, 76, 82, 83, 101]. However, the most common assumption is that failures occur independently and with the same distribution, being the Poisson, renewal, or phase-type (PH) renewal processes the usual arrival processes used to model the fails. These assumptions are unrealistic in practice, since inter-failures are usually correlated. Therefore, there is a need for suitable point processes that properly fit the occurrence of failures.

We will consider an N electrical component framework. These electrical components are generators that are assumed to have the same structure, since they are all built in the same way. The available data are the operational times of each component. After each failure occurs, we assume instant repair times, which we believe to be a good approximation since the repair times are very short. The aim is to determine (estimate) the process that governs the functioning of each component which is the same for all the components. That is, we have N independent samples of inter-failure times related to the same process. A similar data structure has been considered in other works, for example, Fiondella et al. [32] analyzed the system reliability of a system of N components considering correlated component failures (CCF) under a multivariate Bernoulli distribution approach. Der Kiureghian and Song [25] examined the reliability of a system of N two-state components by decomposing it into a number of subsystems. The decomposition is combined with a linear programming formulation, which allows for the computation of the system probability bounds for the reliability analysis. More recently, Montoro-Cazorla and Pérez-Ocón [72] studied the failure times of a cold standby N components system assuming that the occurrences of shocks and repairs follow a stationary MAP processes. It is also remarkable that the introduced data framework can be found in several other contexts where recurrent events occur, such as medicine, where the components are the patients and the occurrences are the recurrent times of an infection [1, 65]; or aeronautical engineering, where the failures of the air conditioning system of each member of a fleet of airplanes are considered [85].

In this chapter, we consider a real data set with the previously described framework, provided by Iberdrola, which is a private electrical utility company that supplies energy to Spain. After analyzing the Iberdrola data set, we were able to verify that the operational times associated with each generator cannot be assumed to be independent nor identically distributed. Considering these properties and the fact that the components are assumed to have the same structure, it is reasonable to assume that the failure times are derived from the same process. It will be shown that an appropriate model to fit the operational times is the non-stationary two-state MAP, whose matricial formulation enables the calculation of significant quantities and probabilities of interest associated to the reliability of the system, such as the cumulative distribution of the inter-failure times, the probabilities and expected number of failures at a specific time t. Recently, the stationary MAP has been considered in a reliability context to model the arrival of shocks in shock and wear systems, see [17, 69, 70, 71, 84]. To the best of our knowledge, and unlike its stationary counterpart (see [12]), inference for the non-stationary MAP has not been considered before. Therefore, the main objective of this chapter is to develop a parameter estimation approach for fitting the non-stationary two-state *MAP* (MAP_2) to N sequences of operational times (inter-failure times).

This chapter is organized as follows. Section 3.1 introduces the framework of the data and the definition. Special emphasis is put on reviewing some performance measures of the system as the cumulative distribution function and moments of the times between failures, correlation between the times of consecutive failures or the probability of N failures at a specific time t. Section 3.2 describes the considered parameter estimation methodology, namely, a moments matching approach, which is illustrated with a simulation study in Section 3.3. In Section 3.4 we apply the proposed algorithm to real failure times of N electrical components in the Iberdrola dataset. Finally, in Section 3.5 we provide some conclusions. Most of the results of this chapter can be found in Rodríguez, Lillo and Ramírez-Cobo [98].

3.1 The model

We consider an N electrical component framework, each of them subject to failures. After each failure, the component is instantly repaired, and we record the consecutive operational times of each component. The aim is to estimate the performance of the system on base on N real sequences of such operational times, $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(N)}$, where

$$\begin{aligned} \mathbf{t}^{(1)} &= \left(t_1^{(1)}, t_2^{(1)}, \ldots, t_{n_1}^{(1)}\right), \\ \mathbf{t}^{(2)} &= \left(t_1^{(2)}, t_2^{(2)}, \ldots, t_{n_2}^{(2)}\right), \\ &\vdots \\ \mathbf{t}^{(N)} &= \left(t_1^{(N)}, t_2^{(N)}, \ldots, t_{n_N}^{(N)}\right), \end{aligned} \tag{3.1}$$

where n_i denotes the size of the sample $\mathbf{t}^{(i)}$, for $i = 1, ..., N$.

We assume that the N components are identical and the sequences of operational times $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(N)}$, are mutually independent.

Let T_k be the random variable representing the operational time between the $(k-1)$-th and k-th failure of any given component. Unlike classical model assumptions, we cannot assume that the random variables $\{T_k\}_{k\geq 1}$ are uncorrelated, and then, we do not consider them independent. Also, the random variables $\{T_k\}_{k\geq 1}$ are not necessarily identically distributed.

In Chapter 2 we proved that the non-stationary MAP_2 is not unique, and provided a canonical, unique representation of the non-stationary MAP_2 in Theorem 2.3. For the sake of simplicity, we rewrite them here. When $\gamma > 0$:

$$\boldsymbol{\alpha} = (\alpha,\ 1-\alpha), \quad D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} -x-y & 0 \\ v & -u-v \end{pmatrix}. \tag{3.2}$$

On the contrary, if $\gamma \leq 0$, the canonical representation is given by

$$\boldsymbol{\alpha} = (\alpha,\ 1-\alpha), \quad D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & -x-y \\ -u-v & v \end{pmatrix}, \tag{3.3}$$

where $u \leq x < 0$, $y, v \geq 0$, $x + y \leq 0$ and $u + v \leq 0$.

3.1.1 Performance of the system.

Most of the descriptors necessary to measure the performance of the system have been introduced in the Section 1.2.3.1, defined in (1.5), (1.6), (1.9) and (1.10), which we proceed to remember: the time between the $(k-1)$-th and the k-th failures, T_k, is phased-type distributed with representation $\{\boldsymbol{\alpha}_k, D_0\}$, where

$$\boldsymbol{\alpha}_k = \boldsymbol{\alpha}\left(P^{\star}\right)^{k-1},$$

which implies that the inter-failure times are not identically distributed. Then the cumulative distribution function and moments of T_k are given by

$$F_{T_k}(t) = 1 - \boldsymbol{\alpha}_k e^{D_0 t}\mathbf{e},$$

$$\mu_{k,m} = E\left(T_k^m\right) = m!\boldsymbol{\alpha}_k\left(-D_0\right)^{-m}\mathbf{e},$$

for $m = 1, 2, \ldots$ and where $\mathbf{e}$ represents a vector of ones. We point out that, given the operational times as in (3.1), then each column of data $t_k^{(1)}, t_k^{(2)}, \ldots, t_k^{(N)}$ are observations of the random variable T_k: they represent times between the $(k-1)$-th and k-th failures of the first, second,... and $N-$th component. Therefore, the corresponding correlation function between T_k and T_{k+l}, which shows that the inter-failure times are not independent, is given in the next result.

Lemma 3.1. *The correlation between T_k and T_{k+l}, for $l = 2, 3, \ldots$ is given by*

$$corr\left(T_k, T_{k+l}\right) = \frac{\left[\boldsymbol{\alpha}_k\left(-D_0\right)^{-1}\left(P^{\star}\right)^l\left(-D_0\right)^{-1}\mathbf{e}\right] - \left[\boldsymbol{\alpha}_k\left(-D_0\right)^{-1}\mathbf{e}\right]\left[\boldsymbol{\alpha}_{k+l}\left(-D_0\right)^{-1}\mathbf{e}\right]}{\sqrt{2\boldsymbol{\alpha}_k\left(-D_0\right)^{-2}\mathbf{e} - \left(\boldsymbol{\alpha}_k\left(-D_0\right)^{-1}\mathbf{e}\right)^2}\sqrt{2\boldsymbol{\alpha}_{k+l}\left(-D_0\right)^{-2}\mathbf{e} - \left(\boldsymbol{\alpha}_{k+l}\left(-D_0\right)^{-1}\mathbf{e}\right)^2}}. \tag{3.4}$$

See Appendix 3.A for the proof. In addition, Eq. (3.4) shows that the inter-failure times are not independent, which is another interesting property of the *MAP*.

Concerning the counting process $\{N(t),\ t \geq 0\}$, recall that the probability of n failures at time t is given by,

$$P\left(N(t) = n \mid N(0) = 0\right) = \boldsymbol{\alpha} P(n,t)\mathbf{e},$$

where the probability of n failures in the interval $(0,t]$ is given by the matrix $P(n,t)$, defined in (1.8). And the expected number of failures at time t, $E\left(N(t) \mid N(0) = 0\right)$, is computed from the first factorial moment of the counting process,

$$M_1(t) = \sum_{n=0}^{\infty} nP(n,t).$$

A final remark concerning the difference between the non-stationary MAP_2 and its stationary version needs to be made at this point. The non-stationary MAP_2, represented by $\{\boldsymbol{\alpha}, D_0, D_1\}$, converges to its steady-state version, $\{\boldsymbol{\phi}, D_0, D_1\}$, when t increases (or $k \to \infty$). In this situation, the probability vector $\boldsymbol{\alpha}_k$, converges to $\boldsymbol{\phi}$, the stationary probability vector of $P^{\star}$. In particular, in the stationary version of the process the sequence of times between failures $\{T_k\}_{k\geq 1}$ are identically distributed as a random variable $T \sim PH\left(\boldsymbol{\phi}, D_0\right)$, and hence if the steady-state MAP_2 were considered instead, we could not assume in our model that the random variables $\{T_k\}_{k\geq 1}$ are not identically distributed.

3.2 Parameter estimation algorithm

In this section we present a method for estimating the parameters of the non-stationary MAP_2 described in (3.2)-(3.3) given a set of operational times $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(N)}$ as in (3.1). As it was showed in Chapter 2, the non-stationary MAP_2 is characterized by the set of moments $\{\mu_{1,1},\ \mu_{1,2},\ \mu_{1,3},\ \mu_{2,1},\ \mu_{3,1}\}$, where $\mu_{k,m} = \mu_{k,m}\left(\alpha, x, y, u, v\right)$ is defined in (1.6). Therefore, it seems reasonable to define a moment matching estimation approach where the population moments $\mu_{k,m}$ are matched by their empirical counterparts $\overline{\mu_{k,m}}$, computed as

$$\overline{\mu_{k,m}} = \frac{1}{N}\sum_{i=1}^{N}\left(t_k^{(i)}\right)^m. \tag{3.5}$$

This leads to solve the nonlinear system of equations defined by

$$\begin{aligned} \mu_{1,m}\left(\alpha, x, y, u, v\right) &= \overline{\mu_{1,m}},\ m = 1,2,3, \\ \mu_{k,1}\left(\alpha, x, y, u, v\right) &= \overline{\mu_{k,1}},\ k = 2,3. \end{aligned} \tag{3.6}$$

The previous problem may not have a feasible solution when real data are considered. Therefore, in order to obtain an estimate, we follow Carrizosa and Ramírez (2012) [12], and seek the model parameters fulfilling (3.6) as much as possible. Specifically, we propose to solve the optimization problem

$$(P)\begin{cases} \min & \delta_\tau(\alpha, x, y, u, v) \\ s.t. & x, u \leq 0, \\ & y, v \geq 0, \\ & -x-y \geq 0, \\ & -u-v \geq 0, \\ & 0 \leq \alpha \leq 1, \end{cases}$$

where $\delta_\tau(\alpha, x, y, u, v)$ is the objective function given by

$$\begin{aligned} \delta_\tau(\alpha, x, y, u, v) &= \tau\Bigg\{ \left(\frac{\mu_{1,1}-\overline{\mu_{1,1}}}{\overline{\mu_{1,1}}}\right)^2 + \left(\frac{\mu_{1,2}-\overline{\mu_{1,2}}}{\overline{\mu_{1,2}}}\right)^2 + \left(\frac{\mu_{1,3}-\overline{\mu_{1,3}}}{\overline{\mu_{1,3}}}\right)^2 + \\ &\quad + \left(\frac{\mu_{2,1}-\overline{\mu_{2,1}}}{\overline{\mu_{2,1}}}\right)^2 + \left(\frac{\mu_{3,1}-\overline{\mu_{3,1}}}{\overline{\mu_{3,1}}}\right)^2 \Bigg\}, \end{aligned}$$

and $\tau > 0$ is a penalty parameter that needs to be tuned. In our experience, we set $\tau = 1$, which seems to perform well in practice. Clearly, $(\hat{\alpha}, \hat{x}, \hat{y}, \hat{u}, \hat{v})$ solves (3.6) if and only if it is an optimal solution of (P), whose optimal value is 0. Problem (P) is solved in practice with MATLAB using the command `fmincon`, which finds the minimum of a constrained nonlinear multivariable function. Its default optimization methodology is based on a trust-region-reflective algorithm (see [19], for instance).

Given the sample $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(N)}$, problem (P) needs to be solved twice, one per each of the two canonical representations (3.2) and (3.3). The estimated parameters under the model with highest log-likelihood are selected, where the likelihood function for each sequence $\mathbf{t}^{(i)}$, $i = 1, \ldots, N$, is given by

$$f(\mathbf{t}^{(i)}|D_0, D_1) = \boldsymbol{\alpha} e^{D_0 t_1^{(i)}} D_1 e^{D_0 t_2^{(i)}} D_1 \ldots e^{D_0 t_{n_i}^{(i)}} D_1 \mathbf{e}.$$

Therefore, from the assumption over the model, the log-likelihood of the sample is

$$\log f(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(N)}|D_0, D_1) = \sum_{i=1}^{N} \log f(\mathbf{t}^{(i)}|D_0, D_1). \tag{3.7}$$

Once the parameters of the non-stationary MAP_2 are inferred, it is straightforward to obtain estimations of the quantities of interest concerning the counting process, namely, the probability of N failures in $(0, t]$ and the expected number of failures at time t.

The numerical computation of the matrices $P(n, t)$ is based on the uniformization method addressed in Neuts and Li [77], where the $P(n, t)$ matrices can be written as

$$P(n,t) = \sum_{r=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} V(n,r), \tag{3.8}$$

where λ is a constant such that $\lambda \geq \max_j\{-(D_0)_{jj}\}$ and $V(n, r)$ are 2×2 matrices recursively computed as (see [77]):

S.1 Define
$$C_0 = I + \frac{D_0}{\lambda} \qquad \text{and} \qquad C_1 = \frac{D_1}{\lambda}.$$

S.2 Find the smallest index N for which $\sum_{n=N+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} \leq \epsilon$, where ϵ is a fixed tolerance parameter.

S.3 For $n = 0$, $V(0, 0) = I$, $P(0, t) \leftarrow V(0, 0)e^{-\lambda t}$.

S.4 For $n \geq 1$, $V(n, 0) \leftarrow 0$, $P(n, t) \leftarrow 0$.

S.5 For $1 \leq k \leq N$ and $0 \leq i \leq k$,

$$\begin{aligned} V(i,k) &= \sum_{j=max\{0,i-(k-1)\}}^{min\{i,1\}} V(i-j,k-1)C_j, \\ P(i,t) &\leftarrow P(i,t) + V(i,k)e^{-\lambda t}\frac{(\lambda t)^k}{k!}. \end{aligned}$$

3.3 Numerical results

In this section, we perform a a couple of simulational studies in order to clarify the proposed estimation approach.

Example 3.1. *Consider a sample of operational times as in (3.1), simulated from the MAP_2 defined by*

$$\boldsymbol{\alpha} = (0.9252, 0.0748), \quad D_0 = \begin{pmatrix} -0.683 & 0.0026 \\ 0 & -34.6904 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0.6804 \\ 34.5586 & 0.1318 \end{pmatrix}, \tag{3.9}$$

where $N = 1000$ *and* $n_1 = \ldots = n_{1000} = 100$.

The theoretical and sample moments are given by

$$\begin{aligned}(\mu_{1,1},\ \mu_{1,2},\ \mu_{1,3},\ \mu_{2,1},\ \mu_{3,1}) &= (1.3927, 2.0370, 2.9823, 0.1053, 1.3829),\\ (\overline{\mu_{1,1}},\ \overline{\mu_{1,2}},\ \overline{\mu_{1,3}},\ \overline{\mu_{2,1}},\ \overline{\mu_{3,1}}) &= (0.1154, 0.1441, 0.1949, 1.4262, 0.1193).\end{aligned}$$

The solution to problem (P) *under the first canonical form* (3.2) *was*

$$\hat{\boldsymbol{\alpha}}^1 = (0.9219, 0.0781),\quad \hat{D}_0^1 = \begin{pmatrix}-0.73921 & 0.7392\\ 0 & -102.6784\end{pmatrix},\quad \hat{D}_1^1 = \begin{pmatrix}0 & 0\\ 9.0049 & 93.6735\end{pmatrix},$$

with 95% confidence intervals given by

$$CI_{95\%}(\alpha) = (0, 1),\quad CI_{95\%}(x) = (-126.7766, -0.7252),\quad CI_{95\%}(y) = (0.0001, 57.8303),$$
$$CI_{95\%}(u) = (-122.4563, -0.7230),\quad CI_{95\%}(v) = (0.0001, 65.5339),$$

and estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (0.1154, 0.1441, 0.1949, 0.1284, 0.1284),$$

and objective function equal to $\delta_\tau^1 = 0.8339$.

On the other hand, the estimated MAP_2 *parameters under the canonical form* (3.3) *were*

$$\hat{\boldsymbol{\alpha}}^2 = (0.9365, 0.0635),\quad \hat{D}_0^2 = \begin{pmatrix}-0.6801 & 0.0020\\ 0 & -42.1618\end{pmatrix},\quad \hat{D}_1^2 = \begin{pmatrix}0 & 0.6781\\ 42.1617 & 0.0001\end{pmatrix},$$

with 95% confidence intervals given by

$$CI_{95\%}(\alpha) = (0.01, 0.99),\quad CI_{95\%}(x) = (-130.4271, -0.6800),\quad CI_{95\%}(y) = (0.0001, 52.7877),$$
$$CI_{95\%}(u) = (-136.0214, -0.6801),\quad CI_{95\%}(v) = (0.0001, 106.6294),$$

and estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (0.1156, 0.1379, 0.2020, 1.3789, 0.1193),$$

and objective function $\delta_\tau^2 = 0.0043$. *In order to select the appropriate canonical from, the log-likelihoods as in* (3.7) *were computed:*

$$\log f(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(1000)} | \hat{D}_0^1, \hat{D}_1^1) = -86085.2730,\quad \log f(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(1000)} | \hat{D}_0^2, \hat{D}_1^2) = -8126.3579,$$

which provides strong evidence in favor of the estimate $\{\hat{\alpha}^2, \hat{D}_0^2, \hat{D}_1^2\}$*. We calculated the estimations of (3.9) under different values of* τ*.*

Finally, in order to test the influence of the tuning parameter, τ*, consider Table 3.1, which shows the estimations of the parameters under different values of* τ*,* $\tau \in \{1, 10, 50, 100\}$*. As can be deduced, in this case and for this sample, the choice of* τ *is not relevant, with the exception of* $\tau = 50$*, which provides a solution with permuted states.*

τ	$\hat{\alpha}^2$	$\hat{D}_0^2$	$\hat{D}_1^2$
1	$(0.9365, 0.0635)$	$\begin{pmatrix} -0.6801 & 0.0020 \\ 0 & -42.1618 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.6781 \\ 42.1617 & 0.0001 \end{pmatrix}$
10	$(0.9365, 0.0635)$	$\begin{pmatrix} -0.6801 & 0.0020 \\ 0 & -42.1607 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.6781 \\ 42.1606 & 0.0001 \end{pmatrix}$
50	$(0.0635, 0.9365)$	$\begin{pmatrix} -0.42.1609 & 0.1182 \\ 0 & -0.6801 \end{pmatrix}$	$\begin{pmatrix} 0 & 42.0427 \\ 0.68 & 0.0001 \end{pmatrix}$
100	$(0.9365, 0.0635)$	$\begin{pmatrix} -0.6801 & 0.0020 \\ 0 & -42.1607 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.6781 \\ 42.1607 & 0.0001 \end{pmatrix}$

Table 3.1: Point estimates of the model parameters under an assortment of values of τ.

Example 3.2. *In this example,unlike the previous one, we also illustrate the estimated probabilities* $P(n, t)$ *for* $n = 1, 3, 5$*.*

Consider a sample of operational times as in (3.1), simulated from the MAP_2 *defined by*

$$\boldsymbol{\alpha} = (0.3172, 0.6828), \quad D_0 = \begin{pmatrix} -1.6290 & 0.1326 \\ 0 & -5.0987 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 1.4964 \\ 4.6152 & 0.4835 \end{pmatrix}, \quad (3.10)$$

where $N = 1000$ *and* $n_1 = \ldots = n_{1000} = 200$*.*

The theoretical and sample moments are given by

$$\begin{aligned} (\mu_{1,1},\ \mu_{1,2},\ \mu_{1,3},\ \mu_{2,1},\ \mu_{3,1}) &= (0.3337, 0.1499, 0.0812, 0.4743, 0.3574), \\ (\overline{\mu_{1,1}},\ \overline{\mu_{1,2}},\ \overline{\mu_{1,3}},\ \overline{\mu_{2,1}},\ \overline{\mu_{3,1}}) &= (0.4731, 0.2653, 0.1602, 0.3641, 0.4590). \end{aligned}$$

The solution to problem (P) *under the first canonical form (3.2) was*

$$\hat{\boldsymbol{\alpha}}^1 = (0.4202, 0.5798), \quad \hat{D}_0^1 = \begin{pmatrix} -1.5882 & 1.5882 \\ 0 & -4.7951 \end{pmatrix}, \quad \hat{D}_1^1 = \begin{pmatrix} 0 & 0 \\ 1.4641 & 3.3311 \end{pmatrix},$$

with 95% confidence intervals given by

$$CI_{95\%}(\alpha) = (0, 1), \quad CI_{95\%}(x) = (-94.2810, -1.6122), \quad I_{95\%}(y) = (1.1068, 67.6450),$$
$$CI_{95\%}(u) = (-92.9675, -1.6911), \quad CI_{95\%}(v) = (0.0001, 69.0836),$$

and estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (0.4733, 0.2653, 0.1602, 0.6470, 0.6470),$$

and objective function equal to $\delta_\tau^1 = 0.0262$.

On the other hand, the estimated MAP_2 *parameters under the canonical form (3.3) were*

$$\hat{\boldsymbol{\alpha}}^2 = (0.3819, 0.6181), \quad \hat{D}_0^2 = \begin{pmatrix} -1.5885 & 0.0552 \\ 0 & -4.8025 \end{pmatrix}, \quad \hat{D}_1^2 = \begin{pmatrix} 0 & 1.5334 \\ 4.3226 & 0.4699 \end{pmatrix},$$

with 95% confidence intervals given by

$$CI_{95\%}(\alpha) = (0.1949, 0.99), \quad CI_{95\%}(x) = (-98.5473, -1.5887), \quad CI_{95\%}(y) = (0.0543, 57.4087),$$
$$CI_{95\%}(u) = (-107.2629, -1.7151), \quad CI_{95\%}(v) = (0.0001, 51.2046),$$

and estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (0.4731, 0.2653, 0.1602, 0.3641, 0.4590),$$

and objective function $\delta_\tau^2 = 1.0464 \times 10^{-9}$. *In order to select the appropriate canonical from, the log-likelihoods as in (3.7) were computed:*

$$\log f(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(1000)} | \hat{D}_0^1, \hat{D}_1^1) = -4.1510 \times 10^5, \quad \log f(\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(1000)} | \hat{D}_0^2, \hat{D}_1^2) = -1.0013 \times 10^{19},$$

which provides strong evidence in favor of the estimate $\{\hat{\alpha}^2, \hat{D}_0^2, \hat{D}_1^2\}$.

To illustrate the estimated probabilities of having $n = 1, 3, 5$ *failures over time, we calculate the matrices* $P(n, t)$ *implementing the procedure introduced in Section 3.2, with a tolerance parameter fixed as* $\epsilon = 0.001$. *Figure 3.1 illustrates the values of the probabilities (1.9). The solid and dashed lines show the theoretical probabilities of the original* MAP_2, $\{\boldsymbol{\alpha}, D_0, D_1\}$, *given in (3.10), and the estimated* MAP_2*,* $\{\hat{\alpha}^2, \hat{D}_0^2, \hat{D}_1^2\}$ *respectively, and the dotted line represents the empirical probabilities of the simulated data. We can observe that the estimating procedure performs reasonably providing an acceptable fit for the given data.*

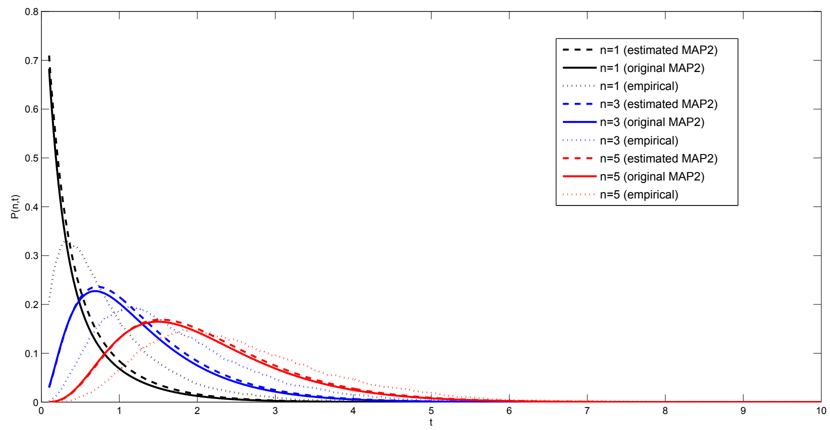

Figure 3.1: Estimated probabilities $P(N(t) = n \mid N(0) = 0)$ for $n = 1, 3, 5$ and $t > 0$.

Example 3.3. *In this example we discuss the computational cost of the proposed algorithm. To such aim, consider Table 3.2, which depicts the computational times in seconds for* $N \in \{50, 100, 500, 1000\}$ *and* $n \in \{50, 100, 150, 200\}$*. Note that for* $N = \{50, 100\}$ *the execution time tends to decrease as* n *increases, whereas for* $N = \{500, 1000\}$*, the execution time tends to increase as* n *increases.*

		N			
		50	100	500	1000
n	50	141,7047	128,1592	143,5250	169,3486
	100	132,1727	131,1166	123,7231	187,7061
	150	131,8070	130,4978	143,5805	187,6100
	200	131,9233	117,4214	161,2899	193,1425

Table 3.2: Computational times for (N, n), where $N \in \{50, 100, 500, 1000\}$ and $n \in \{50, 100, 150, 200\}$.

To construct Table 3.2, we selected 600 *randomly chosen starting points for which the stopping criteria were established as follows:*

1. `MaxFunEvals`. *The bound on the number of function evaluations is fixed to* 500.

2. `MaxIter`. *The bound on the number of solver iterations is set to* 400.

3. `TolFun`. *The lower bound on the change in the value of the objective function during a step has been set to* 1.000×10^{-6}.

4. `TolX`. *The lower bound on the stepsize has been set to* 1.000×10^{-6}.

5. `TolCon`. *The upper bound on the magnitude of any constraint functions has been fixed as* 1.000×10^{-6}.

The computational times (and all calculations) were calculated using MATLAB® version 7.1.0.246 (R14) on a PC with processor Intel Core i7-3537U, 2.5 GHz.

3.4 Illustration with a real data set

In this section we first show a numerical example of parameter estimation for a non-stationary MAP_2 from a real data set provided by the Spanish private electrical utility company, Iberdrola. Second, we highlight important quantities regarding the counting process for the estimated non-stationary MAP_2, and third, we compare the obtained results with those obtained by its stationary counterpart.

3.4.1 Data description

The electrical failure times provided by the company Iberdrola, are collected from an electrical framework composed of 926 components that supply energy. The times in which the components fail are recorded. These components are generators that are assumed to be built in the same way; that is, they are structurally the same. If an electrical component or generator fails it is repaired and restarted. The repair times are considered negligible in comparison with the operational times.

The data presents an structure as in (3.1), where

$$\begin{aligned} &N = 926, \\ &\min\{n_1, \ldots, n_N\} = 1, \\ &\max\{n_1, \ldots, n_N\} = 42, \end{aligned} \tag{3.11}$$

where n_i, for $i = 1, ..., N$, denotes the number of observed operational times of the i-th component.

Consider the random variables $\{T_k\}_{k\geq 1}$ with sample values given by $\{t_k^{(i)}\}_{i\in\{1,...,N\}}$. Note from (3.11) that not all $\{T_k\}_{k\geq 1}$ will be characterized by a total of $N = 926$ sample values, and in particular the random variables $\{T_k\}_{k\geq 1}$, for k *close* to 42 will be characterized by few samples, a fact that affects to the computation of empirical quantities, as will be seen. In this work, components $i \in \{1, \ldots, 926\}$ for which $n_i < 3$ (16 out of 926 components) will not be considered in the estimation approach described in Section 3.2 since the corresponding operational times for computing the sample moments (3.5) are missing.

The 926 components are considered to be equal, since the company states they are built with the same structure. Therefore the assumption over the model described in Section 3.1 is fulfilled.

The correlation matrix of the random vector $(T_1, T_2, \ldots T_{42})$ was estimated. As commented in Section 3.4.1 the sequences of sample values $\{t_k^{(i)}\}_{i\in\{1,...,N\}}$, $k \in \{1, \ldots, 42\}$, may possess a length less than 926. In our experiment, only samples of length larger than 30 (a total of 25 sequences out of 42) will be considered. This led to a final estimated correlation matrix of size 25×25 (that is, the variables $T_{26}, \ldots, T_{42}$ were not taken into account for being characterized by a low number of sample data). We found that a total of 32 (out of 300) pairs (T_k, T_l), $k, l \in \{1, \ldots, 25\}$, $k < l$, presented a correlation coefficient ranging in $[0.25, 0.7194]$. In addition, 11 (out of 300) pairs had a correlation coefficient which ranged in $[-0.3266, -0.25]$. These results provide evidence enough to consider a correlation structure. Therefore, it is necessary to use a model to fit the data that allows dependent failures.

Finally, a Kolmogorov-Smirnov (K-S) test was performed to determine if the samples $\{t_k^{(i)}\}_{i\in\{1,...,N\}}$ and $\{t_l^{(i)}\}_{i\in\{1,...,N\}}$, for $k, l \in \{1, \ldots, 42\}$ and $k < l$, are drawn from the same underlying continuous populations. As in the previous discussion, only samples of length larger than 30 were considered for testing the equality in distribution. Our findings show that the equality in distribution is rejected for the 52% of pairs of such samples, which implies that the inter-failure times cannot be consider identically distributed nor independent.

3.4.2 Performance estimation under the non-stationary MAP_2

We show in this section how to fit the non-stationary MAP_2 to the statistical pattern of the data. The sample moments needed for estimating the model are

$$(\overline{\mu_{1,1}},\ \overline{\mu_{1,2}},\ \overline{\mu_{1,3}},\ \overline{\mu_{2,1}},\ \overline{\mu_{3,1}}) = (79.226, 7.478\times 10^3, 7.2911\times 10^3, 69.0582, 67.5977).$$

The inference approach described in the previous section provided the following estimates of the non-stationary MAP_2 in the first canonical form:

$$\hat{\boldsymbol{\alpha}}^1 = (0.4608, 0.5392), \quad \hat{D}_0^1 = \begin{pmatrix} -0.2394 & 0.1345 \\ 0 & -0.0104 \end{pmatrix}, \quad \hat{D}_1^1 = \begin{pmatrix} 0.1049 & 0 \\ 0.0067 & 0.0037 \end{pmatrix},$$

with estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (78.8950, 7.535\times 10^3, 7.2633\times 10^3, 69.0864, 67.5712),$$

and objective function equal to $\delta_\tau^1 = 9.0324\times 10^{-5}$.

On the other hand, the estimated MAP_2 parameters under the second canonical form were

$$\hat{\boldsymbol{\alpha}}^2 = (0.8207, 0.1793), \quad \hat{D}_0^2 = \begin{pmatrix} -0.0104 & 0.0104 \\ 0 & -16.5378 \end{pmatrix}, \quad \hat{D}_1^2 = \begin{pmatrix} 0 & 0 \\ 11.7651 & 4.7727 \end{pmatrix},$$

with estimated moments given by

$$(\widehat{\mu_{1,1}},\ \widehat{\mu_{1,2}},\ \widehat{\mu_{1,3}},\ \widehat{\mu_{2,1}},\ \widehat{\mu_{3,1}}) = (78.7930, 7.5583\times 10^3, 7.2513\times 10^3, 68.3119, 68.3119),$$

and objective function $\delta_\tau^2 = 4.0324\times 10^{-4}$. The log-likelihoods as in (3.7) were

$$\log f(\mathbf{t}^{(1)},\ldots,\mathbf{t}^{(N)}|\hat{D}_0^1,\hat{D}_1^1) = -5.3790\times 10^4, \quad \log f(\mathbf{t}^{(1)},\ldots,\mathbf{t}^{(N)}|\hat{D}_0^2,\hat{D}_1^2) = -5.7335\times 10^4,$$

which provides evidence in favor of the estimate $\{\hat{\alpha}^1, \hat{D}_0^1, \hat{D}_1^1\}$.

Dashed line in Figure 3.1 depicts the estimated distribution functions of the random variables T_1, T_2 and T_3 under the non-stationary MAP_2 (see 1.5), while the solid line shows the empirical cdfs. It can be seen how the fit to the cdfs of the consecutive times between failures looks reasonable. In order to validate the results, a χ^2-goodness-of-fit was performed. The obtained $p-$values, which are shown in Figure 3.1, imply that the null hypotheses cannot be rejected at at 0.001 significance level.

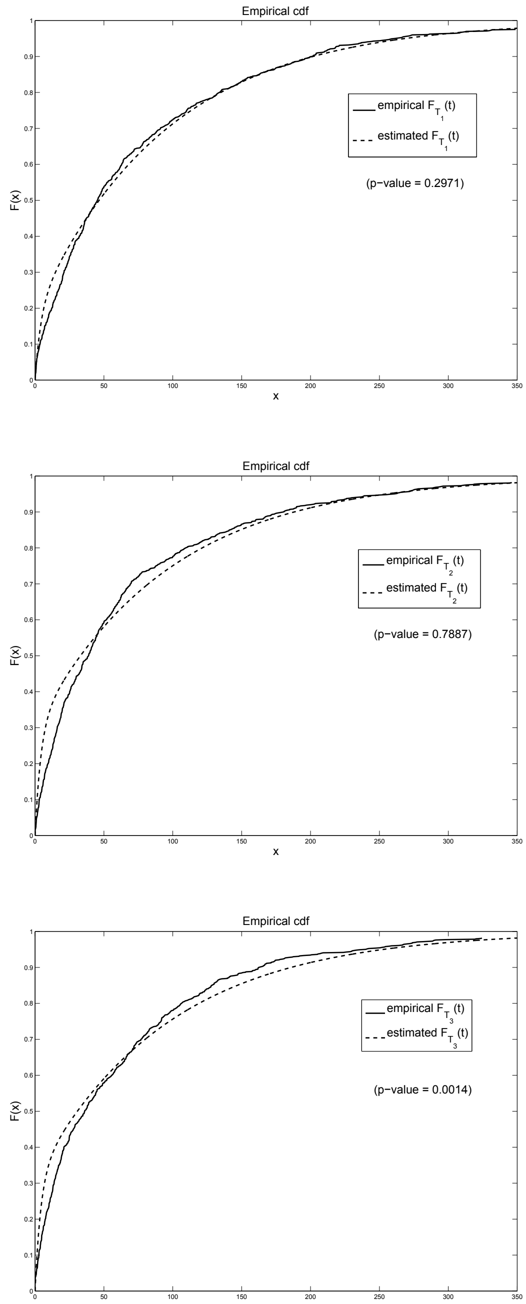

Figure 3.1: Estimated cdf (dashed line) under the non-stationary MAP_2 versus the empirical cdf (solid line) of the random variables T_1 (time until the first failure, top panel), T_2 (time between the first and second failure, central panel), and T_3 (time between the second and third failure, bottom panel).

We focus now on the quantities of interest associated to the counting process. Given the selected estimate $\{\hat{\alpha}^1, \hat{D}_0^1, \hat{D}_1^1\}$, the algorithm to calculate the matrices $P(n,t)$ introduced in Section 3.2 is implemented, with a tolerance parameter fixed as $\epsilon = 0.001$ (as in Example 3.2). Figure 3.2 shows the values of the probabilities (1.9) for an assortment of values (n,t), $n \in \mathbb{N}$ and $t > 0$.

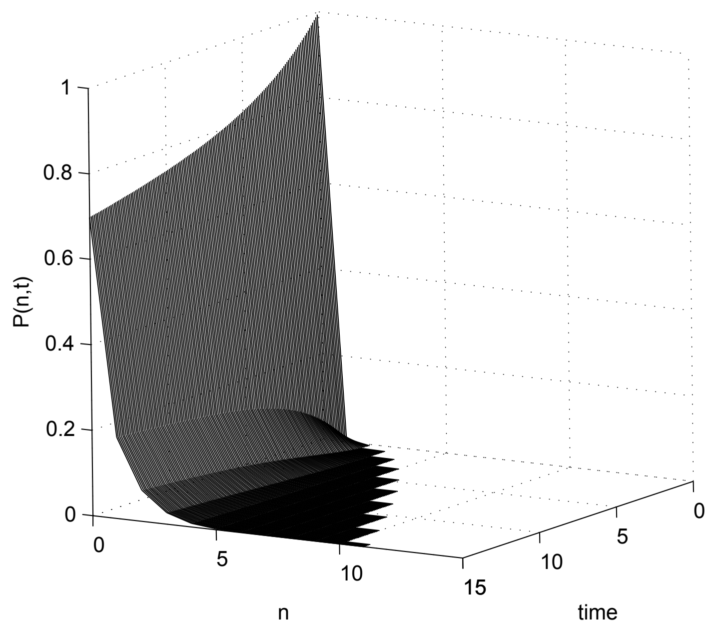

Figure 3.2: Probabilities $P\left(N(t) = n \mid N(0) = 0\right)$ for $n \in \mathbb{N}$ and $t > 0$.

Figure 3.3 shows the probabilities of having $n = 1, 2, 3, 4, 5$ failures over time. It can be observed that the probability of having 1 failure is high at first, but then decreases slowly over time. However, the probability of having 2, 3, 4 or 5 failures slowly increases. Finally, Figure 3.4 depicts the expected number of failures at time t, $E\left(N(t) \mid N(0) = 0\right)$, for a sequences of times $t > 0$, and as expected, such value escalates as the time progresses.

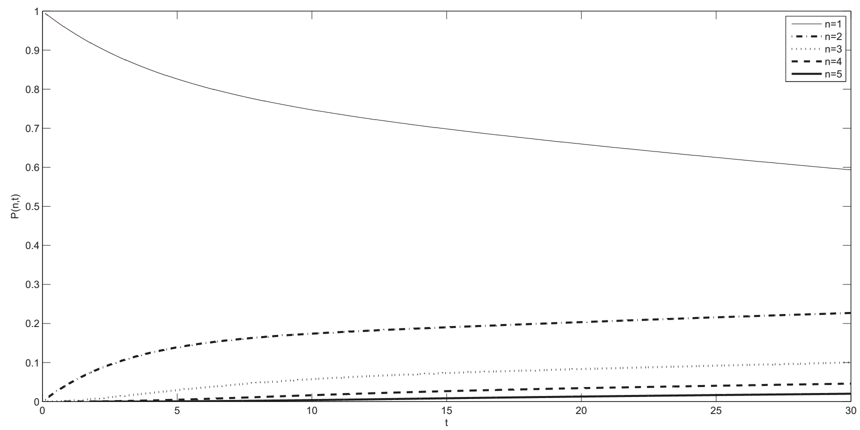

Figure 3.3: Probabilities $P\left(N(t)=n \mid N(0)=0\right)$ for $n=1,2,3,4,5$ and $t>0$.

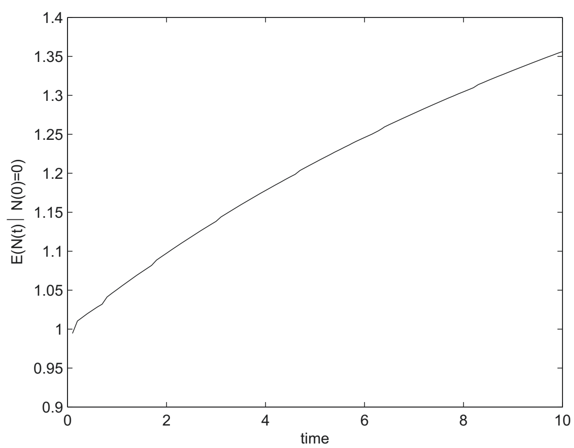

Figure 3.4: Expected number of failures at time t, $E\left(N(t) \mid N(0)=0\right)$.

3.4.3 Comparison with the stationary $\boldsymbol{MAP_2}$

In this section we compare some of the previous results obtained under the non-stationary MAP_2 with those provided by its stationary counterpart. As commented at the end of Section 3.1.1, the stationary version of the MAP_2 turns out a less versatile process since the sequence of random variables $\{T_k\}_{k\geq 0}$ are assumed to be equally distributed as a random variable $T \sim PH\left(\boldsymbol{\phi}, D_0\right)$. From the discussion at the end of Section 3.4.1 about equality in distribution, it is thus expected that the non-stationary MAP_2 outperforms its stationary version when fitting the considered real data set.

The stationary MAP_2 is completely characterized by the sequence $\{\mu_1, \mu_2, \mu_3, \rho_1\}$, where $\mu_m = E(T^m)$, $m \in \mathbb{N}$, and ρ_1 is the first-lag auto-correlation coefficient, see Bodrog et al. [7]. Therefore, the approach described in Carrizosa and Ramírez-Cobo [12], based on the matching of the population moments $\{\mu_1, \mu_2, \mu_3, \rho_1\}$ by their empirical estimates is applied. A remark concerning the computation of the sample moments $\{\overline{\mu_1}, \overline{\mu_2}, \overline{\mu_3}, \overline{\rho_1}\}$ from a sample as in (3.1) is needed at this point. Since in this case the variables $\{T_k\}_{k\geq 0}$ are assumed to be equally distributed, it is reasonable to estimate μ_m as

$$\overline{\mu_m} = \frac{\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left(t_i^{(j)}\right)^m}{\sum_{j=1}^{N} n_j},$$

that is, by considering all observations $t_i^{(j)}$ generated from the same random variable T. In the case of the considered real data set, the values

$$\overline{\mu_1} = 56.8406,\ \overline{\mu_2} = 8.2932 \times 10^3,\ \overline{\mu_3} = 1.956 \times 10^6$$

were obtained.

The computation of an empirical estimate of the first-lag auto-correlation coefficient is not so straightforward. On one hand, a data structure as in (3.1) leads to N different estimates of ρ_1, and on the other hand, for those components $i \in \{1, \ldots, N\}$ for which n_i is *small* the computation of an estimate of ρ_1 is not possible. In this work, two different estimates for ρ_1 were considered. First, denote by $\hat{\rho}_1$ the estimated first-lag auto-correlation coefficient of the sample $\mathbf{t}^{(i)}$ for which $n_i = \max\{n_1, \ldots, n_N\}$ (that is, the estimated ρ_1 is that of the longest sample $\mathbf{t}^{(i)}$). The second estimate of ρ_1, noted as $\tilde{\rho}_1$, is given by median of the auto-correlation coefficients of the 1% of the longest samples $\mathbf{t}^{(i)}$, $i \in \{1, \ldots, N\}$. In our real data base, the values

$$\hat{\rho}_1 = -0.2293, \quad \tilde{\rho}_1 = 0.1778$$

were obtained.

Under the choice $\overline{\rho_1} = \hat{\rho}_1$, the fitted stationary MAP_2 led to the estimated moments

$$(\widehat{\mu_1}, \widehat{\mu_2}, \widehat{\mu_3}, \widehat{\rho_1}) = \left(56.8403,\ 8.2931 \times 10^3,\ 1.9597 \times 10^6,\ 0\right)$$

Similarly, under the choice $\overline{\rho_1} = \tilde{\rho}_1$, the results were

$$(\widehat{\mu_1}, \widehat{\mu_2}, \widehat{\mu_3}, \widehat{\rho_1}) = \left(56.09,\ 8.5231 \times 10^3,\ 1.9343 \times 10^6,\ 0.1776\right).$$

Note that under the first election of $\overline{\rho_1}$, the sample moments $\overline{\mu_m}$, $m = 1, 2, 3$, are correctly matched; however, the estimate of the empirical first-lag auto-correlation coefficient is poor. In the second case, the empirical first-lag auto-correlation coefficient is well fitted at the expense of a poorer estimation of the sample moments. In order to analyze the cost of considering the random variables $\{T_k\}_{k\geq 0}$ as equally distributed, see Figure 3.5, which depicts the empirical distribution function of the random variables T_1, T_2 and T_3 (in solid line) and the estimated counterparts by the stationary MAP_2 under the two possible choices of $\overline{\rho_1}$, $\hat{\rho}_1$ and $\tilde{\rho}_1$ (in dotted and dashed lines, respectively). The three panels illustrate how the stationary MAP_2 performs poorly than the non-stationary version (see Figure 3.1), especially in the case of T_1, which clearly follows a different distribution from that of the stationary T.

3.5 Chapter summary

In this chapter we have presented a moments method estimation procedure to fit the non-stationary second-order MAP to sequences of operational times of N electrical components that are structurally equal. Unlike previous approaches, here the operational times are considered to be dependent and not identically distributed, an assumption which is realistic in practice. We have also provided the correlation function for the presented data framework. From the estimated parameters of the model, a number of key performance measures regarding the counting process, as the probability of N failures or the expected number of failures at time t, are inferred. The suitability of our approach is illustrated by a simulated example and a real data set example, with operational times provided by the Spanish electrical group Iberdrola. The example highlights the superiority, in terms of modeling, of the non-stationary MAP over its stationary version.

Appendix

3.A Proof of formula (3.4)

The correlation between T_k and T_{k+l}, is given by the following formula

$$corr(T_k, T_{k+l}) = \frac{E(T_k T_{k+l}) - E(T_k)E(T_{k+l})}{s.d.(T_k)s.d.(T_{k+l})}, \tag{3.12}$$

where $s.d.$ stands for standard deviation. We know from (1.6), that

$$\begin{aligned} E\left(T_k\right) &= \boldsymbol{\alpha}\left(P^{\star}\right)^{k-1}\left(-D_0\right)^{-1}\mathbf{e}, \\ E\left(T_k^2\right) &= 2!\boldsymbol{\alpha}\left(P^{\star}\right)^{k-1}\left(-D_0\right)^{-2}\mathbf{e} \end{aligned} \tag{3.13}$$

then

$$\begin{aligned} s.d(T_k) &= \sqrt{E\left(T_k^2\right) - E\left(T_k\right)^2} \\ &= \sqrt{2\boldsymbol{\alpha}\left(P^{\star}\right)^{k-1}\left(-D_0\right)^{-2}\mathbf{e} - \left(\boldsymbol{\alpha}\left(P^{\star}\right)^{k-1}\left(-D_0\right)^{-1}\mathbf{e}\right)^2}. \end{aligned} \tag{3.14}$$

To calculate $E(T_kT_{k+l})$, we follow page 83 of Rodríguez [96], where a close form expression for the expectation of the product of two different phase type random variables is provided. Then, using the parameters that define T_k, we obtain,

$$E(T_kT_{k+l}) = \boldsymbol{\alpha}(P^{\star})^{k-1}\left(-D_0\right)^{-1}\left(P^{\star}\right)^{l}\left(-D_0\right)^{-1}\mathbf{e}. \tag{3.15}$$

Finally, substituting (3.13), (3.14) and (3.15) in (3.12), we obtain (3.4).

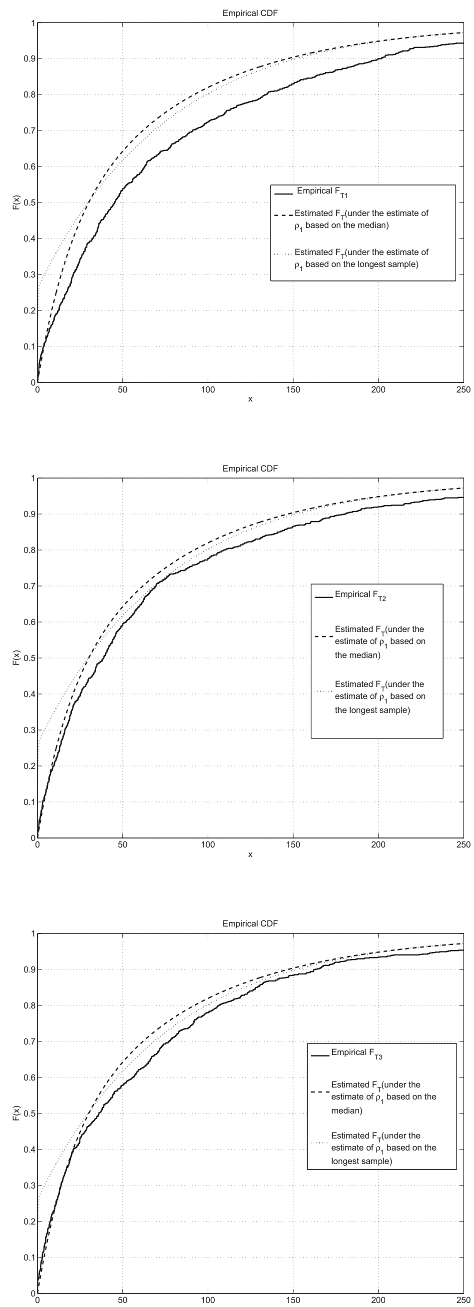

Figure 3.5: Empirical cdf of the random variables T_1, T_2 and T_3 (in solid line in the top, central and bottom panels, respectively) and the estimated counterparts by the stationary MAP_2 under the two possible choices of $\overline{\rho_1}$, $\hat{\rho}_1$ and $\tilde{\rho}_1$ (in dotted and dashed lines, respectively).

Chapter 4

Non-identifiability of the two-state *BMAP*

The Batch Markovian Arrival Process (*BMAP*) was introduced in Section 1.2.3.2 as a generalization of the *MAP* process that allows for correlated batch event occurrences. Let us recall it, the $BMAP_m(k)$ is a doubly stochastic process $\{J(t), N(t)\}$, where $J(t)$ represents an irreducible, continuous, Markov process with state space $\mathcal{S} = \{1, \ldots, m\}$ and $N(t)$ is a counting process where the transitions from (i, n) to $(j, n + k_0)$ correspond to batch arrivals of size $k_0 \leq k$, $i, j \in \mathcal{S}$. The $BMAP_m(k)$ can be characterized by rate matrices $\mathcal{B} = \{D_0, D_1, \ldots, D_k\}$ with matrices elements given by

$$\begin{aligned}(D_0)_{ii} &= -\lambda_i, \quad i = 1, 2,\\(D_0)_{ij} &= \lambda_i p_{ij0}, \quad i, j = 1, 2, \ i \neq j,\\(D_l)_{ij} &= \lambda_i p_{ijl}, \quad i, j = 1, 2, \ 1 \leq l \leq k.\end{aligned}$$

*BMAP*s are highly-parametrized models where, in practice, only inter-event times and batch sizes are usually observed. Therefore, as discussed in Section 1.3 and Chapter 2 for the *MAP*, which are *BMAP*s with arrivals of size 1, it is common to encounter identifiability problems. In the context of *BMAP*s such a property may be formulated along the lines of Rydén [94] or Ramírez-Cobo et al. [87]. Specifically, if T_n and B_n represent the time between the $(n-1)$-th and n-th event occurrences, and the batch arrival size of the n-th event occurrence in a *BMAP* noted by $\mathcal{B}$, then $\mathcal{B}$ is said to be non-identifiable if there exits a differently parametrized *BMAP*, noted as $\widetilde{\mathcal{B}}$, such that

$$(T_1, \ldots, T_n, B_1, \ldots, B_n) \stackrel{d}{=} \left(\widetilde{T}_1, \ldots, \widetilde{T}_n, \widetilde{B}_1, \ldots, \widetilde{B}_n\right), \quad \text{for all } n \geq 1, \tag{4.1}$$

where $\stackrel{d}{=}$ denotes equality of joint distributions, and $\widetilde{T}_n$ and $\widetilde{B}_n$ represents the inter-event times and batch sizes of the *BMAP* noted as $\widetilde{\mathcal{B}}$.

While performance analysis for models incorporating *BMAP*s is a well developed area, less progress has been made on statistical estimation for such models (for estimation of the *MAP*, see Kriege and Buchholz [52] and the references given there and Section 1.3). The study of the identifiability is crucial when estimation of the process parameters is to be considered. In particular, the non-identifiability of a process has serious negative consequences: the likelihood function has infinitely more global maxima and 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. However, neither of the previous studies considered the inclusion of batch arrivals, we refer the reader to Section 1.3 and the introduction of Chapter 2 for more detail in this topic.

In this chapter we address the problem of identifiability of the stationary version of the two-state *BMAP*, *BMAP*$_2(k)$, where k is the maximum batch arrival size. Under the assumptions that the inter-event times and batch sizes are observed, we prove that the stationary *BMAP*$_2(k)$ is a non-identifiable process for $k \geq 2$. We also provide a method to obtain an equivalent *BMAP*$_2(k)$ to a given one, considering the k *BMAP*$_2(2)$s obtained as combinations of the rate matrices defining the given *BMAP*$_2(k)$, for $k \geq 2$.

This chapter is organized as follows. Section 4.1 contains some novel results concerning the batches size distribution. In Section 4.2 our main result is stated and proved, namely, the *BMAP*$_2(2)$ is a non-identifiable process. Section 4.3 generalizes the previous result to the case $k \geq 3$. Finally, in Section 4.4 we provide conclusions to this chapter. Most of the results of this chapter can be found in Rodríguez, Lillo and Ramírez-Cobo [97].

4.1 Distributional properties of the batch arrival size

This section generalizes previous results on identifiability of *MAP*s to the case where, not only the inter-event times but also batch arrivals are observed. As will be shown, our approach involves looking at the distributional properties of the stationary batch arrival size, B.

First, the probability function of B is given by

$$P(B = l) = \boldsymbol{\phi}(-D_0)^{-1} D_l \mathbf{e}, \quad \text{for} \quad l = 1, \ldots, k,$$

where $(-D_0)^{-1} D_l$ is the transition probability matrix of arrivals of size equal to l.

from which the moments of B are obtained as

$$E[B^n] = \boldsymbol{\phi}(-D_0)^{-1}D_n^{\star}\mathbf{e},$$

where $D_n^{\star} = \sum_{l=1}^{k} l^n D_l$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_{BMAP}$ is the stationary probability vector defined in (1.14) as

$$\boldsymbol{\phi} = (\boldsymbol{\pi} D\mathbf{e})^{-1}\boldsymbol{\pi} D,$$

where $\boldsymbol{\pi} = \boldsymbol{\pi}_{BMAP}$ is the stationary probability vector of the underlying Markov process $J(t)$.

Second, the joint moment generating function of the inter-event times and batch sizes in the stationary version of the process, $(\boldsymbol{T}, \boldsymbol{B})$ where $\boldsymbol{T} = (T_1, \ldots, T_n)$ and $\boldsymbol{B} = (B_1, \ldots, B_n)$, is established by the next Lemma.

Lemma 4.1. *The moment generating function of the n first inter-event times and batch sizes of a stationary BMAP$_2$(k) is given by*

$$f^{*}_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s}, \boldsymbol{z}) = \boldsymbol{\phi}(s_1 I - D_0)^{-1}\xi(z_1)\ldots(s_n I - D_0)^{-1}\xi(z_n)\mathbf{e}, \tag{4.2}$$

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

See Appendix 4.A for the proof. The auto-correlation function of the batch sizes is obtained as an immediate result of the previous Lemma.

Lemma 4.2. *Let B_n represent the n-th batch event occurrence size in the stationary version of the BMAP$_2$(k). Then, the auto-correlation function, $\rho(B_1, B_n)$ is given by*

$$\rho(B_1, B_n) = \frac{\boldsymbol{\phi}(-D_0)^{-1}D_1^{\star}\left[(-D_0)^{-1}D\right]^{n-2}(-D_0)^{-1}D_1^{\star}\mathbf{e} - \mu_B^2}{\sigma_B^2}, \tag{4.3}$$

where $\mu_B = E(B)$ and $\sigma_B^2 = Var(B)$.

See Appendix 4.B for the proof. The auto-correlation function of the time between the $(n-1)$-th and n-th event occurrences, in the stationary version $\rho(T_1, T_n)$, is given by,

$$\rho(T_1, T_n) = \frac{\mu_T \boldsymbol{\pi}\left[(-D_0)^{-1}D\right]^{n-1}(-D_0)^{-1}\mathbf{e} - \mu_T^2}{\sigma_T^2}, \tag{4.4}$$

where $\mu_T = E(T)$ and $\sigma_T^2 = V(T)$ are the mean and variance of the inter-event times in its stationary version. A detailed analysis for the correlation structure of (4.3) and (4.4) is given in Chapter 5.

We should point out here that the lemmas previously stated also hold for the case of the BMAP$_m$(k), where $m \geq 2$.

4.2 Non-identifiability of the $\boldsymbol{BMAP}_2(2)$

As commented previously, the identifiability of a process is of critical importance when inference is considered since the lack of a unique representation implies infinite solutions and possibly non-convergence of the typical maximum likelihood approaches. In this section we prove the non-identifiability of the stationary $BMAP_2(2)$, or in other words, the existence of a differently parametrized representation of the process.

4.2.1 Preliminaries

Consider the non-identifiability definition given by (4.1). From now on, the notation $\mathcal{B} \equiv \widetilde{\mathcal{B}}$ will be used to represent equivalence between two given $BMAP_2(k)$s with representations $\mathcal{B}$ and $\widetilde{\mathcal{B}}$. Note that the equality in distribution (4.1) is equivalent to the equality of the moment generating functions defined in (4.2),

$$f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s}, \boldsymbol{z}) = f^*_{\widetilde{\boldsymbol{T}};\widetilde{\boldsymbol{B}}}(\mathbf{s}, \boldsymbol{z}), \tag{4.5}$$

for all $\mathbf{s} = (s_1, \ldots, s_n)$, $\boldsymbol{z} = (z_1, \ldots, z_n)$ and all $n \geq 1$. As will be seen, the proof for the non-identifiability of the $BMAP_2(2)$ consists on the existence of infinite solutions to the system of equations given by (4.1) in terms of the Laplace transforms.

From now on, a stationary $BMAP_2(2)$ will be represented by $\mathcal{B} = \{D_0, D_1, D_2\}$ where

$$D_0 = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} w & m \\ v & q \end{pmatrix}, \quad D_2 = \begin{pmatrix} n & -x-y-w-m-n \\ t & -r-u-v-q-t \end{pmatrix}, \tag{4.6}$$

where without loss of generality it is assumed that $x \leq u$. According to (1.11),

$$x = -\lambda_1, \quad y = \lambda_1 p_{120}, \quad w = \lambda_1 p_{111}, \quad m = \lambda_1 p_{121}, \quad n = \lambda_1 p_{112},$$

$$r = \lambda_2 p_{210}, \quad u = -\lambda_2, \quad v = \lambda_2 p_{211}, \quad q = \lambda_2 p_{221}, \quad t = \lambda_2 p_{212}.$$

The stationary probability distribution $\boldsymbol{\phi} = (\phi, 1-\phi)$ is computed from (1.14) in terms of the model parameters as

$$\phi = \frac{rn + rw - xt - xv}{ux - yr + rn + rw - xt - xv - yt - yv + un + uw}.$$

Given a $BMAP_2(2)$ defined by $\mathcal{B} = \{D_0, D_1, D_2\}$, then it is clear that $\mathcal{M} = \{R_0 = D_0, R_1 = D_1 + D_2\}$ defines a MAP_2, where ocurrences of batches 1 or 2 are considered as being equal, that is

$$R_0 = D_0 = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \quad R_1 = D_1 + D_2 = \begin{pmatrix} w+n & -x-y-w-n \\ v+t & -r-u-v-t \end{pmatrix}. \tag{4.7}$$

From now on, a MAP_2 as in (4.7) constructed from a $BMAP_2(2)$ defined by $\mathcal{B} = \{D_0, D_1, D_2\}$, will be called underlying MAP_2 to a given $BMAP_2(2)$. Assume we look for a $BMAP_2(2)$ defined by $\widetilde{\mathcal{B}} = \{\widetilde{D}_0, \widetilde{D}_1, \widetilde{D}_2\}$ that is equivalent to $\mathcal{B}$ according to Definition 4.1. In particular the equality

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

must hold, and therefore the underlying MAP_2 defined by $\widetilde{\mathcal{M}} = \{\widetilde{R}_0 = \widetilde{D}_0, \widetilde{R}_1 = \widetilde{D}_1 + \widetilde{D}_2\}$ must be equivalent to $\mathcal{M}$, according to Definition 1 in Ramírez-Cobo et al. [87]. Our proof for the non-identifiability of the $BMAP_2(2)$ will rely on this result about the equivalence of the underlying MAP_2s. In particular, in order for two $BMAP_2(2)$s, $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, to be equivalent, the equality in distribution

$$T_1 \stackrel{d}{=} \widetilde{T}_1,$$

must hold, or equivalently,

$$\frac{s(\alpha+\gamma)+\eta}{s^2+s\upsilon+\eta} = \frac{s(\tilde{\alpha}+\tilde{\gamma})+\tilde{\eta}}{s^2+s\tilde{\upsilon}+\tilde{\eta}} \quad \text{for all } s, \tag{4.8}$$

where α, γ, η, υ (respectively $\tilde{\alpha}$, $\tilde{\gamma}$, $\tilde{\eta}$, $\tilde{\upsilon}$) are given by

$$\begin{aligned} \alpha &= \phi(m+w-v-q)+(v+q), \\ \gamma &= \phi(-m-w-x-y+q+u+v+r)-(q+u+v+r), \\ \eta &= xu-yr, \\ \upsilon &= -u-x. \end{aligned} \tag{4.9}$$

Define the value of τ as

$$\tau = \eta + (\alpha+\gamma)(\alpha+\gamma-\upsilon). \tag{4.10}$$

Ramírez-Cobo and Lillo [91] determined that (4.8) holds if and only if

$$\begin{pmatrix} -1 & 0 & 0 \\ -\upsilon & \alpha+\gamma & -1 \\ -\eta & \eta & \alpha+\gamma-\upsilon \end{pmatrix} \begin{pmatrix} \tilde{\alpha}+\tilde{\gamma} \\ \tilde{\upsilon} \\ \tilde{\eta} \end{pmatrix} = \begin{pmatrix} -\alpha-\gamma \\ -\eta \\ 0 \end{pmatrix} \tag{4.11}$$

where τ is the determinant of the coefficient matrix associated to the linear system (4.11). Then, if $\tau \neq 0$, the system (4.11) possesses a unique solution given by

$$\widetilde{\alpha} + \widetilde{\gamma} = \alpha + \gamma, \quad \widetilde{\eta} = \eta, \quad \widetilde{\upsilon} = \upsilon.$$

Otherwise, we have that $\eta = (\alpha+\gamma)(\upsilon-\alpha-\gamma)$, and there are infinite solutions, specified by

$$\widetilde{\alpha} + \widetilde{\gamma} = \alpha + \gamma, \quad \widetilde{\eta} = (\alpha+\gamma)(\widetilde{\upsilon}-\alpha-\gamma), \text{ for } \widetilde{\upsilon} > 0.$$

In Ramírez-Cobo et al. [87], a procedure to construct an equivalent MAP_2 to a differently parametrized given MAP_2 is provided, under the assumption that the MAP_2 parameters satisfy $\tau \neq 0$. In this work we focus on $BMAP_2(2)$s for which their underlying MAP_2s also satisfy $\tau \neq 0$. It can be proven that if $\tau = 0$, then the inter-event times $(T_1, T_2, \ldots, T_n, \ldots)$ are independent and identically PH-distributed random variables with representation $\{\boldsymbol{\phi}, D_0\}$. Since the identifiability of the PH-distribution has been widely studied in the literature (see O´ Cinneide [79] or Rydén [94]) and the independence of the inter-event times may be seen not interesting from a modeling viewpoint, then for the sake of abbreviation we do not consider such cases.

Consider now the equality (4.1) for $n = 1$, which from Lemma 4.1 is equivalent to

$$\boldsymbol{\phi}(sI - D_0)^{-1}\xi(z)\mathbf{e} = \widetilde{\boldsymbol{\phi}}(sI - \widetilde{D}_0)^{-1}\widetilde{\xi}(z)\mathbf{e},$$

or alternatively, to

$$\frac{z(s\alpha+\beta)+z^2(s\gamma-\beta+\eta)}{s^2+s\upsilon+\eta} = \frac{z(s\widetilde{\alpha}+\widetilde{\beta})+z^2(s\widetilde{\gamma}-\widetilde{\beta}+\widetilde{\eta})}{s^2+s\widetilde{\upsilon}+\widetilde{\eta}}, \tag{4.12}$$

where α, γ, η, υ (respectively $\widetilde{\alpha}$, $\widetilde{\gamma}$, $\widetilde{\eta}$, $\widetilde{\upsilon}$) are given by (4.9), and β (respectively $\widetilde{\beta}$) is given by

$$\beta = \phi(-uw + yv - um + yq - rw + xv - rm + xq) + (rw - xv + rm - xq). \tag{4.13}$$

Note that since D_0 is invertible, then necessarily $\eta \neq 0$. Define next the following three expressions:

$$\begin{aligned} C_1 &= \alpha\eta - \beta(\alpha+\gamma), \\ C_2 &= \beta + \alpha(\alpha+\gamma-\upsilon), \\ C_3 &= \eta - \beta + \gamma(\gamma-\upsilon+\alpha). \end{aligned}$$

It can be proven that the system of equations defined by (4.12) possesses a unique solution given by

$$\widetilde{\alpha} = \alpha, \quad \widetilde{\beta} = \beta, \quad \widetilde{\gamma} = \gamma, \quad \widetilde{\eta} = \eta, \quad \widetilde{\upsilon} = \upsilon, \tag{4.14}$$

if and only if at least one of the C_is is different from zero, for $i = 1, 2, 3$.

Otherwise, when $C_1 = C_2 = C_3 = 0$, an infinite number of solutions to (4.12) are found:

$$\alpha = \widetilde{\alpha}, \qquad \widetilde{\beta} = \beta - \alpha(\upsilon - \widetilde{\upsilon}), \qquad \gamma = \widetilde{\gamma}, \qquad \widetilde{\eta} = \eta - (\alpha + \gamma)(\upsilon - \widetilde{\upsilon}), \quad \text{for} \quad \widetilde{\upsilon} > 0.$$

Consider now (4.1) for $n = 2$. We proceed analogously to the case $n = 1$ to find that (4.5) becomes

$$\phi(s_1 I - D_0)^{-1}\xi(z_1)(s_2 I - D_0)^{-1}\xi(z_2)\mathbf{e} = \widetilde{\phi}(s_1 I - \widetilde{D}_0)^{-1}\widetilde{\xi}(z_1)(s_2 I - \widetilde{D}_0)^{-1}\widetilde{\xi}(z_2)\mathbf{e}, \tag{4.15}$$

where

$$\begin{aligned}
&\phi(s_1 I - D_0)^{-1}\xi(z_1)(s_2 I - D_0)^{-1}\xi(z_2)\mathbf{e} = \\
&\frac{z_1 z_2(s_1\delta_1 + s_2\delta_2 + s_1 s_2\delta_3 + \delta_4) + z_1 z_2^2(s_1(\alpha\eta - \delta_1) + s_2\delta_5 + s_1 s_2\delta_6)}{(s_1^2 + s_2^2)\eta + s_1^2 s_2^2 + (s_1^2 s_2 + s_1 s_2^2)\upsilon + (s_1 + s_2)\eta\upsilon + s_1 s_2\upsilon^2 + \eta^2} + \\
&\frac{z_1 z_2^2(\beta\eta - \delta_4) + z_1^2 z_2(s_1\delta_7 + s_2\delta_8 + s_1 s_2\delta_9 + \delta_{10})}{(s_1^2 + s_2^2)\eta + s_1^2 s_2^2 + (s_1^2 s_2 + s_1 s_2^2)\upsilon + (s_1 + s_2)\eta\upsilon + s_1 s_2\upsilon^2 + \eta^2} + \\
&\frac{z_1^2 z_2^2(s_1(\eta\gamma - \delta_7) + s_2\delta_{11} + s_1 s_2\delta_{12} + \eta^2 - \eta\beta - \delta_{10})}{(s_1^2 + s_2^2)\eta + s_1^2 s_2^2 + (s_1^2 s_2 + s_1 s_2^2)\upsilon + (s_1 + s_2)\eta\upsilon + s_1 s_2\upsilon^2 + \eta^2},
\end{aligned}$$

(respectively the right-hand side of (4.15)) where δ_i for $i = 1, \ldots, 12$ are given by

$$\begin{aligned}
\delta_1 &= \beta(q - m) + (\phi(w - v - q + m) + v)(vy - mu + qy - uw) \\
&\quad -m(qx - mr - rw + vx), \\
\delta_2 &= \beta(q + w) + (\phi(x + y - r - u) + (r - x))(mv - qw), \\
\delta_3 &= \alpha(q + w) + mv - qw, \\
\delta_4 &= \beta(mr - qx - uw + vy) + \eta(mv - qw), \\
\delta_5 &= \phi((r + u)(mr + mu - qx - qy) + (x + y)(-vx - vy + wr + wu)) \\
&\quad +(r + u)(qx - mr) + (x + y)(vx - wr) - \delta_2, \\
\delta_6 &= \phi((r + u)(q - m) + (x + y)(v - w)) - q(r + u) - v(x + y) - \delta_3, \qquad (4.16)\\
\delta_7 &= (\gamma - n)(rm + rw - xq - xv) - \beta(t - n) + t(-mu + qy - uw + vy), \\
\delta_8 &= (\eta - \beta)(q + v) + (\phi(nr + nu - tx - ty) + (tx - nr))(q + v - m - w), \\
\delta_9 &= \gamma(q + v) + (\phi(t - n) - t)(q + v - m - w),
\end{aligned}$$

$$\begin{aligned}
\delta_{10} &= \eta(mr - qx - vx + rw + mt + tw - nq - nv) \\
&\quad -\beta(mr - qx - vx + rw + nr + nu - tx - ty), \\
\delta_{11} &= (\phi(-nr - nu + tx + ty) + (rn - tx))(r + u - x - y) \\
&\quad +(r + u)(\beta - \eta) - \delta_8, \\
\delta_{12} &= -\gamma(r + u) + (r + u - x - y)(\phi(n - t) + t) - \delta_9,
\end{aligned}$$

respectively, $\widetilde{\delta}_i$ for $i = 1, \ldots, 12$. Then, it is tedious but straightforward to prove that if $C_1 \neq 0$ or $C_2 \neq 0$ or $C_3 \neq 0$, then (4.14) leads to (for a detailed proof see Appendix 4.C)

$$\widetilde{\delta}_i = \delta_i, \quad \text{for } i = 1, \ldots, 12,$$

and therefore, in any of these cases the solution to the equality in distributions (4.1) for $n = 1, 2$ is

$$\widetilde{\alpha} = \alpha, \quad \widetilde{\beta} = \beta, \quad \widetilde{\gamma} = \gamma, \quad \widetilde{\eta} = \eta, \quad \widetilde{\upsilon} = \upsilon, \quad \widetilde{\delta}_i = \delta_i, \quad \text{for } i = 1, \ldots, 12. \tag{4.17}$$

Given a known $BMAP_2(2)$ defined by $\mathcal{B} = \{D_0, D_1, D_2\}$ (or alternatively by $x, y, \ldots, t$) as in (4.6), then the set of equations given by (4.17) provides the solutions $\widetilde{\mathcal{B}} = \{\widetilde{D}_0, \widetilde{D}_1, \widetilde{D}_1\}$ (alternatively $\tilde{x}, \tilde{y}, \ldots, \tilde{t}$) of a differently parametrized $BMAP_2(2)$, such that (4.5) holds for $n = 1$ and $n = 2$. In Ramírez-Cobo et al. [87] the analogous equations for the MAP_2 are solved. Then, it is proved there that the (infinite) solutions also satisfy the equations for $n \geq 3$ and therefore, the MAP_2 is concluded to be non-identifiable. However, due to the complexity of the set of expressions (4.9), (4.13) and (4.16), this approach was unfeasible in practice, since it was not possible to (symbolically) obtain the values of $\widetilde{\mathcal{B}} = \{\widetilde{D}_0, \widetilde{D}_1, \widetilde{D}_1\}$ $(\tilde{x}, \tilde{y}, \ldots, \tilde{t})$ that solve the system (4.17) via standard symbolic packages as Maple or Matlab. In consequence, a different approach for solving the identifiability problem of the $BMAP_2(2)$ needs to be considered. The proof of the main result in next section shows such a procedure.

Finally, after extensive simulational experiments, it has been checked that the most frequent $BMAP_2(2)$s are those for which $C_1 \cdot C_2 \cdot C_3 \neq 0$. In what follows we assume that a given $BMAP_2(2)$ satisfies this property together with the previously discussed $\tau \neq 0$. Note that $\tau = C_2 + C_3$, and therefore the assumption that $\tau \neq 0$ combined with $C_1 \cdot C_2 \cdot C_3 \neq 0$ leads to $C_1 \cdot C_2 \cdot C_3 \neq 0$ and $C_2 \neq -C_3$. The remaining cases (that is, $C_1 \cdot C_2 \cdot C_3 = 0$, but still $\tau \neq 0$) are considered in the Appendix 4.D.

4.2.2 Main result

The next result provides the solutions to (4.17).

Proposition 4.1. *Let $\mathcal{B}$ be a BMAP$_2(2)$ as in (4.6) with underlying MAP$_2$, $\mathcal{M}$, as in (4.7). Assume that*

A1. $C_1 \cdot C_2 \cdot C_3 \neq 0$.

A2. $\mathcal{M}$ *satisfies* $\tau \neq 0$.

Let $\tilde{u} < 0$ and $\tilde{r} > 0$, and let $\tilde{x}(\tilde{u},\tilde{r})$, $\tilde{y}(\tilde{u},\tilde{r})$, $\tilde{v}(\tilde{u},\tilde{r})$, $\tilde{w}(\tilde{u},\tilde{r})$, $\tilde{m}(\tilde{u},\tilde{r})$, $\tilde{q}(\tilde{u},\tilde{r})$, $\tilde{n}(\tilde{u},\tilde{r})$ and $\tilde{t}(\tilde{u},\tilde{r})$ be defined as

$$
\begin{aligned}
\tilde{x}(\tilde{u},\tilde{r}) &= -\tilde{u} + x + u, \\
\tilde{y}(\tilde{u},\tilde{r}) &= -\frac{(\tilde{u}^2 - \tilde{u}x - \tilde{u}u + xu - ry)}{\tilde{r}}, \\
\tilde{q}(\tilde{u},\tilde{r}) &= h_1(\widetilde{\phi}, \tilde{x}, \tilde{y}, D_0, D_1, D_2), \\
\tilde{m}(\tilde{u},\tilde{r}) &= h_2(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2), \\
\tilde{n}(\tilde{u},\tilde{r}) &= h_3(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2), \\
\tilde{t}(\tilde{u},\tilde{r}) &= h_4(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2), \\
\tilde{w}(\tilde{u},\tilde{r}) &= \Bigg(\frac{xun + ryt - \tilde{u}xw + ryv + \tilde{u}xv + \tilde{u}uv - ryn + u\tilde{r}t - \tilde{r}yv + \tilde{u}\tilde{r}n}{\omega} \\
&\quad + \frac{-\tilde{u}\tilde{r}v - \tilde{r}yt + 2\tilde{r}xu - \tilde{r}xw - \tilde{u}\tilde{r}t + u\tilde{r}v - ryw + \tilde{u}xt - \tilde{u}uw}{\omega} \\
&\quad + \frac{-xu^2 - \tilde{u}xn - r\tilde{r}y + rx\tilde{r} + ruy - rxu + \tilde{u}^2w + \tilde{u}^2n - r\tilde{u}^2}{\omega} \\
&\quad + \frac{-\tilde{u}^2t - \tilde{u}^2v + \tilde{u}^3 + \tilde{r}u^2 - \tilde{r}^2x - 2\tilde{u}^2u + r^2y + \tilde{u}u^2 - 2\tilde{u}\tilde{r}x}{\omega} \\
&\quad + \frac{-\tilde{r}^2u - 3\tilde{u}\tilde{r}u + 2\tilde{u}xu + 2\tilde{u}^2\tilde{r} + r\tilde{u}u - \tilde{u}\tilde{r}r - \tilde{u}^2x + \tilde{r}^2\tilde{u}}{\omega} \\
&\quad + \frac{\tilde{u}ut + r\tilde{u}x - \tilde{r}xn + xuw - \tilde{u}un + ru\tilde{r} + r\tilde{r}w + r\tilde{r}n}{\omega} \\
&\quad + \frac{-xuv - xut + \tilde{u}\tilde{r}w - ry\tilde{u}}{\omega} \Bigg) - \tilde{n}(\tilde{u},\tilde{r}), \\
\tilde{v}(\tilde{u},\tilde{r}) &= \Bigg(\frac{\tilde{r}(-xv - xt - yv - yt + rw + rn + uw + un - \tilde{r}w - \tilde{r}n + r\tilde{r} + r\tilde{u})}{\omega} \\
&\quad + \frac{\tilde{r}(\tilde{u}u - \tilde{r}^2 - 2\tilde{u}\tilde{r} + u\tilde{r} + \tilde{r}v + \tilde{r}t + \tilde{u}v + \tilde{u}t - \tilde{u}w - \tilde{u}n - \tilde{u}^2)}{\omega} \Bigg) - \tilde{t}(\tilde{u},\tilde{r}),
\end{aligned}
\tag{4.18}
$$

where

$$\omega = -\tilde{u}u - u\tilde{r} + xu + \tilde{u}^2 + 2\tilde{u}\tilde{r} + \tilde{r}^2 - \tilde{u}x - \tilde{r}x - ry. \tag{4.19}$$

Then, the set of values $\{\tilde{u}, \tilde{r}, \tilde{x}(\tilde{u},\tilde{r}), \tilde{y}(\tilde{u},\tilde{r}), \tilde{v}(\tilde{u},\tilde{r}), \tilde{w}(\tilde{u},\tilde{r}), \tilde{m}(\tilde{u},\tilde{r}), \tilde{q}(\tilde{u},\tilde{r}), \tilde{n}(\tilde{u},\tilde{r}), \tilde{t}(\tilde{u},\tilde{r})\}$ *solves the system of equations given by (4.17), for specific values of functions* h_1*,* h_2*,* h_3 *and* h_4 *(see Remark 4.1 below).*

Remark 4.1. *Closed-form expressions for* h_i*,* $i = 1, 2, 3, 4$ *can be found at*

https://sites.google.com/site/joavrc/software

They have been included neither in Proposition 4.1 nor in the Appendix due to their large extension (around 43 pages).

Remark 4.2. *It might be the case that under specific values of* $\tilde{u}$ *and* $\tilde{r}$*, the denominators in (4.18) are equal to zero. However, since the set of solutions is infinite (as many solutions as the number of* $\tilde{u}$ *and* $\tilde{r}$ *satisfying* $\tilde{u} < 0$ *and* $\tilde{r} > 0$*), we will concentrate on those values which do not imply any numerical inconsistency. In other words, it may be assumed that the set of values (4.18) contains an infinite subset of well defined solutions.*

Proof. Since it is not possible to symbolically solve the set of equations (4.17), the next alternative procedure was applied. Consider a fixed $BMAP_2(2)$ with representation $\mathcal{B} = \{D_0, D_1, D_2\}$ as in (4.6), with underlying MAP_2, $\mathcal{M} = \{R_0, R_1\}$ as in (4.7), then the method provided by Theorem 4.1 in Ramírez-Cobo et al. [87] (see Appendix 4.E) allows to calculate an equivalent MAP_2, under the assumption that $\tau \neq 0$, given as:

$$\tilde{R}_0 = \begin{pmatrix} \tilde{x} & \tilde{y} \\ \tilde{r} & \tilde{u} \end{pmatrix}, \quad \tilde{R}_1 = \begin{pmatrix} \tilde{d}_{111} & -\tilde{x} - \tilde{y} - \tilde{d}_{111} \\ \tilde{d}_{211} & -\tilde{r} - \tilde{u} - \tilde{d}_{211} \end{pmatrix}, \tag{4.20}$$

where $\tilde{r}$ and $\tilde{u}$ are free parameters, and $\tilde{x}$, $\tilde{y}$, $\tilde{d}_{111}$ and $\tilde{d}_{211}$ are obtained as

$$\begin{aligned}
\tilde{x}(\tilde{u},\tilde{r}) &= -\tilde{u} + x + u, \\
\tilde{y}(\tilde{u},\tilde{r}) &= -\frac{(\tilde{u}^2 - \tilde{u}x - \tilde{u}u + xu - ry)}{\tilde{r}}, \\
\tilde{d}_{111} &= \Big(\frac{xun + ryt - \tilde{u}xw + ryv + \tilde{u}xv + \tilde{u}uv - ryn + u\tilde{r}t - \tilde{r}yv + \tilde{u}\tilde{r}n - \tilde{u}\tilde{r}v}{\omega} \\
&\quad + \frac{-\tilde{r}yt + 2\tilde{r}xu - \tilde{r}xw - \tilde{u}\tilde{r}t + u\tilde{r}v - ryw + \tilde{u}xt - \tilde{u}uw - xu^2 - \tilde{u}xn}{\omega} \\
&\quad + \frac{-r\tilde{r}y + rx\tilde{r} + ruy - rxu + \tilde{u}^2w + \tilde{u}^2n - r\tilde{u}^2 - \tilde{u}^2t - \tilde{u}^2v + \tilde{u}^3 + \tilde{r}u^2}{\omega}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{-\tilde{r}^2x - 2\tilde{u}^2u + r^2y + \tilde{u}u^2 - 2\tilde{u}\tilde{r}x - \tilde{r}^2u - 3\tilde{u}\tilde{r}u + 2\tilde{u}xu + 2\tilde{u}^2\tilde{r} + \tilde{r}^2\tilde{u}}{\omega}\\
&+\frac{-\tilde{u}\tilde{r}r - \tilde{u}^2x + r\tilde{u}u + \tilde{u}ut + r\tilde{u}x - \tilde{r}xn + xuw - \tilde{u}un + ru\tilde{r} + r\tilde{r}w}{\omega}\\
&+\left.\frac{+r\tilde{r}n - xuv - xut + \tilde{u}\tilde{r}w - ry\tilde{u}}{\omega}\right),\\
\tilde{d}_{211} \;=\; &\left(\frac{\tilde{r}(-xv - xt - yv - yt + rw + rn + uw + un - \tilde{r}w - \tilde{r}n + r\tilde{r} + r\tilde{u})}{\omega}\right.\\
&+\left.\frac{\tilde{r}(\tilde{u}u - \tilde{r}^2 - 2\tilde{u}\tilde{r} + u\tilde{r} + \tilde{r}v + \tilde{r}t + \tilde{u}v + \tilde{u}t - \tilde{u}w - \tilde{u}n - \tilde{u}^2)}{\omega}\right),
\end{aligned}
$$

where ω is defined in (4.19).

We should point out here that given any MAP_2 as in (4.7), then there exists infinite equivalent MAP_2s as in (4.20), each one constructed from a specific choice of a certain parameter ε (see Theorem 4.1 in Ramírez-Cobo et al. [87]). Therefore, $\widetilde{R}_0$ and $\widetilde{R}_1$ are indeed, $\widetilde{R}_0(\varepsilon)$ and $\widetilde{R}_1(\varepsilon)$. Imposing the condition that equivalent $BMAP_2$s must have equivalent underlying MAP_2s leads to

$$
\widetilde{D_0} = \widetilde{R}_0 = \begin{pmatrix} \tilde{x} & \tilde{y} \\ \tilde{r} & \tilde{u} \end{pmatrix}, \quad \widetilde{D_1} = \begin{pmatrix} \tilde{w} & \tilde{m} \\ \tilde{v} & \tilde{q} \end{pmatrix}, \quad \widetilde{D_2} = \begin{pmatrix} \tilde{n} & -\tilde{x} - \tilde{y} - \tilde{w} - \tilde{m} - \tilde{n} \\ \tilde{t} & -\tilde{r} - \tilde{u} - \tilde{v} - \tilde{q} - \tilde{t} \end{pmatrix},
$$

where necessarily $\widetilde{D_1} + \widetilde{D_2} = \widetilde{R}_1$ or in other words, $\tilde{w} = \tilde{d}_{111} - \tilde{n}$ and $\tilde{v} = \tilde{d}_{211} - \tilde{t}$. This approach, which can be seen to reduce the number of unknown variables from 10 to 4, is illustrated as follows

$$
\begin{array}{ccc}
\mathcal{B} = \{D_0, D_1, D_2\} & & \widetilde{\mathcal{B}} = \left\{\widetilde{D_0}(\varepsilon), \widetilde{D_1}(\varepsilon), \widetilde{D_2}(\varepsilon)\right\} \\
\Big\downarrow {\scriptstyle MAP_2} & & \Big\uparrow {\scriptstyle BMAP_2(2)} \\
\mathcal{M} = \{R_0, R_1\} & \xrightarrow{\text{fix } \varepsilon} & \widetilde{\mathcal{M}} = \left\{\widetilde{R}_0(\varepsilon), \widetilde{R}_1(\varepsilon)\right\}
\end{array}
$$

In order to find the remaining unknowns $\tilde{n}$, $\tilde{t}$, $\tilde{m}$ and $\tilde{q}$, the known values are substituted into the next subset of equations of (4.17),

$$
\widetilde{\beta} = \beta, \quad \widetilde{\gamma} = \gamma, \quad \widetilde{\delta}_3 = \delta_3, \quad \widetilde{\delta}_4 = \delta_4, \tag{4.21}
$$

to yield the expressions given by (4.18). Note that equations (4.21) must hold because assumption *A1* implies (4.17). Finally, it is cumbersome but straightforward to check that the solutions in (4.18) also satisfy the rest of equations in (4.17).

□

Remark 4.3. *The set of values in (4.18) solves the equality of Laplace transforms (4.5) for $n = 1, 2$. Or equivalently, given a BMAP$_2$(2) as in (4.6) it allows to compute the values of $\widetilde{\mathcal{B}} = \{\widetilde{D_0}, \widetilde{D_1}, \widetilde{D_2}\}$ such that (4.5) holds for $n = 1, 2$. However, we should point out here that not all the infinite solutions in (4.18) define real BMAP$_2$(2)s, as the next example shows.*

Example 4.1. *Consider the BMAP$_2$(2) defined by*

$$D_0 = \begin{pmatrix} -7.0666 & 0.0779 \\ 0.0047 & -6.9116 \end{pmatrix}, D_1 = \begin{pmatrix} 4.5829 & 0.4523 \\ 1.3993 & 1.5595 \end{pmatrix}, D_2 = \begin{pmatrix} 0.0354 & 1.9181 \\ 3.3994 & 0.5488 \end{pmatrix},$$

whose underlying MAP$_2$ is

$$R_0 = \begin{pmatrix} -7.0666 & 0.0779 \\ 0.0047 & -6.9116 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 4.6183 & 2.3704 \\ 4.7986 & 2.1083 \end{pmatrix}.$$

If the value of $\varepsilon = 0.0018$ is set in the method derived in Ramírez-Cobo et al. [87] to find an equivalent MAP$_2$ then

$$\widetilde{R}_0 = \begin{pmatrix} -7.0649 & 0.0985 \\ 0.0064 & -6.9134 \end{pmatrix}, \quad \widetilde{R}_1 = \begin{pmatrix} 6.4265 & 0.5398 \\ 6.6068 & 0.3001 \end{pmatrix}$$

is obtained. Computing the values of $\tilde{n}$, $\tilde{t}$, $\tilde{m}$ and $\tilde{q}$ as in (4.18), finally leads to

$$\widetilde{D_0} = \begin{pmatrix} -7.0649 & 0.0985 \\ 0.0064 & -6.9134 \end{pmatrix}, \widetilde{D_1} = \begin{pmatrix} 5.1102 & -0.6433 \\ 1.9265 & 1.0323 \end{pmatrix}, \widetilde{D_2} = \begin{pmatrix} 1.3163 & 1.1831 \\ 4.6803 & -0.7322 \end{pmatrix},$$

which is not a real BMAP$_2$(2), since m and $(-r-u-v-q-t)$ are negative. Note that in spite of that, the equations (4.17) still hold.

The previous example motivates the seek for feasible values of the free parameters $\tilde{u}$ and $\tilde{r}$ such that the equations in (4.18) define a real $BMAP_2(2)$. One possibility is to define $\tilde{u} \equiv u - \kappa$ and $\tilde{r} \equiv r + \kappa$, where $\kappa > 0$ is an auxiliary variable. By substituting these values into the set (4.18), we obtain the set $\mathcal{F} = \{\tilde{u}(\kappa), \tilde{r}(\kappa), \tilde{x}(\kappa), \tilde{y}(\kappa), \tilde{v}(\kappa), \tilde{w}(\kappa), \tilde{m}(\kappa), \tilde{q}(\kappa), \tilde{n}(\kappa), \tilde{t}(\kappa)\}$ defined by

$$\begin{aligned}
\widetilde{\phi}(\kappa) &= \frac{(r+\kappa)\phi}{r}, \\
\tilde{x}(\kappa) &= x + \kappa, \\
\tilde{y}(\kappa) &= \frac{-(\kappa^2 + (x-u)\kappa - ry)}{r+\kappa},
\end{aligned}$$

$$\begin{aligned}
\tilde{q}(\kappa) &= \frac{qr - v\kappa}{r}, \\
\tilde{w}(\kappa) &= \frac{wr + v\kappa}{r}, \\
\tilde{v}(\kappa) &= \frac{v(r+\kappa)}{r}, \\
\tilde{n}(\kappa) &= \frac{nr + t\kappa}{r}, \\
\tilde{t}(\kappa) &= \frac{t(r+\kappa)}{r}, \\
\tilde{m}(\kappa) &= \frac{-(v\kappa^2 - r(q-w)\kappa - mr^2)}{r(r+\kappa)}.
\end{aligned} \tag{4.22}$$

Remark 4.4. *Note that if $r = 0$, then a numerical inconsistency is found in (4.22). Indeed, it can be proven that there exists a finite number of combinations of the elements of $\{D_0, D_1, D_2\}$ (where in all of them at least one parameter is equal to zero), under which (4.22) cannot be calculated. However, our findings show that in those cases it is always possible to obtain another parametrization of $\tilde{u}$ and $\tilde{r}$ (different to $\tilde{u} = u - \kappa$ and $\tilde{r} = r + \kappa$), under which the new set $\mathcal{F}$ is well defined. For the sake of abbreviation we will focus on the general case, where all the elements of $\{D_0, D_1, D_2\}$ are assumed to be strictly positive or negative.*

The following Proposition provides the possible values that the auxiliary variable κ may take in such a way that the set $\mathcal{F}$, given by (4.22), defines a real $BMAP_2(2)$.

Proposition 4.2. *Consider a $BMAP_2(2)$ with representation $\mathcal{B}$ as in (4.6), and define*

$$\begin{aligned}
\kappa_1 &= -x, \\
\kappa_2 &= \frac{u-x}{2}, \\
\kappa_3 &= \frac{r(1-\phi)}{\phi}, \\
\kappa_4 &= \frac{rq}{v}, \\
\kappa_5 &= -\frac{r}{t}(r+u+v+q+t), \\
\kappa_6 &= \frac{(u-x) + \sqrt{(x-u)^2 + 4ry}}{2}, \\
\kappa_7 &= \frac{r}{2v}\left[(q-w) + \sqrt{(q-w)^2 + 4vm}\right],
\end{aligned}$$

$$\kappa_8 = -\frac{r}{2t}\left[u+v+q+n+t+r- \sqrt{(u+v+q+n+t+r)^2+4t(-y-x-w-m-n)}\right].$$

Let κ be chosen from

$$0<\kappa<\min\left\{\kappa_1,\kappa_2,\kappa_3,\kappa_4,\kappa_5,\kappa_6,\kappa_7,\kappa_8\right\}, \quad \text{if } x<u, \tag{4.23}$$

$$0<\kappa<\min\left\{\kappa_1,\kappa_3,\kappa_4,\kappa_5,\kappa_6,\kappa_7,\kappa_8\right\}, \quad \text{if } x=u, \tag{4.24}$$

and setting $\tilde{u}\equiv u-\kappa$ and $\tilde{r}\equiv r+\kappa$. Then there exist an infinite number of BMAP$_2$(2)s, $\widetilde{\mathcal{B}}$, given by $\mathcal{F}=\{\tilde{u}(\kappa),\tilde{r}(\kappa),\tilde{x}(\kappa),\tilde{y}(\kappa),\tilde{v}(\kappa),\tilde{w}(\kappa),\tilde{m}(\kappa),\tilde{q}(\kappa),\tilde{n}(\kappa),\tilde{t}(\kappa)\}$, defined in (4.22), such that (4.5) holds.

The proof of Proposition 4.2 follows the lines of that of Theorem 4.1 in Ramírez-Cobo et al. [87]. For the interested reader, see Appendix 4.F. Finally, in order to prove that the stationary *BMAP*$_2$(2) does not have a unique representation, we have to show that the set of feasible solutions (4.22) where κ is defined as in Proposition 4.2, also satisfies the equality of the Laplace transform (4.5) for all $n\geq 3$. This is shown by the next result.

Finally, in order to prove that the stationary *BMAP*$_2$(2) does not have a unique representation, we have to show that the set of feasible solutions (4.22) where κ is defined as in Proposition 4.2, also satisfies the equality of the Laplace transform (4.5) for all $n\geq 3$. This is shown by the next result.

Corollary 4.1. *The stationary BMAP$_2$(2) is not an identifiable process.*

Proof. The proof of Corollary 4.1 is analogous to the proof of Theorem 4.2 in Ramírez-Cobo et al. [87] (therefore the details are omitted), where in this case $\Delta(s)$ is replaced by $\Delta(s,z)$ defined as

$$\Delta(s,z)=(sI-D_0)^{-1}\xi(z),$$

with parametrization

$$\Delta(s,z)=\begin{pmatrix} a(s,z) & b(s,z)\\ c(s,z) & d(s,z)\end{pmatrix}.$$

□

Consider Example 4.1, according to Proposition 4.2 a value of κ needs to be selected from

$$0 < \kappa < min\{7.0666, 0.0775, 0.0023, 0.1573, 0.0007, 0.0052, 0.0008, 0.0039\}$$

Take for example, $\kappa = 5 \times 10^{-4}$, then the obtained $BMAP_2(2)$ is

$$\widetilde{D}_0 = \begin{pmatrix} -7.0661 & 0.0856 \\ 0.0052 & -6.9121 \end{pmatrix}, \widetilde{D}_1 = \begin{pmatrix} 4.7399 & 0.0859 \\ 1.5562 & 1.4026 \end{pmatrix}, \widetilde{D}_2 = \begin{pmatrix} 0.4167 & 1.7380 \\ 3.7807 & 0.1675 \end{pmatrix},$$

which is a real $BMAP_2(2)$ and according to Corollary 4.1 it makes (4.5) hold for all $\mathbf{s}$, $\boldsymbol{z}$ and n. Therefore, $\mathcal{B} \equiv \widetilde{\mathcal{B}}$.

All calculations to prove the non-identifiability of the stationary $BMAP_2(2)$ have been carried out using MATLAB® version 7.1.0.246 (R14). In the spirit of a reproducible research the codes utilized in this paper are available at

https://sites.google.com/site/joavrc/software

4.3 Non-identifiability of the $\boldsymbol{BMAP_2(k)}$ for $k \geq 3$

In Section 4.2, it has been proven that the stationary $BMAP_2(2)$ is a non-identifiable process and a procedure to construct the equivalent representation to a given one has been derived. This section goes further and ensures the non-identifiability for the stationary $BMAP_2(k)$, for all $k \geq 3$. As will be seen, our approach uses the decomposition of a given $BMAP_2(k)$ into k differently parametrized $BMAP_2(2)$s and the construction of the equivalent process is based on the equivalent $BMAP_2(2)$s to the composing $BMAP_2(2)$s.

Consider a $BMAP_2(k)$, for $k \geq 3$, denoted by $\mathcal{G} = \{D_0, D_1, ..., D_k\}$ with stationary distribution $\boldsymbol{\phi}_{\mathcal{G}}$ as in (1.14). Consider the k $BMAP_2(2)$s, obtained as combinations of the rate matrices defining $\mathcal{G}$:

$$\mathcal{B}_1 = \left\{D_0, D_1, \sum_{i=2}^{k} D_i\right\}, \ldots, \mathcal{B}_j = \left\{D_0, \sum_{i=1, i\neq j}^{k} D_i, D_j\right\}, \quad (4.25)$$
$$\ldots, \mathcal{B}_k = \left\{D_0, \sum_{i=1, i\neq k}^{k} D_i, D_k\right\}.$$

In order to apply the results of Section 4.2 we will assume that the underlying and common stationary MAP_2 for all $\mathcal{B}_i$, for $i = 1, ..., k$, $\mathcal{M} = \left\{D_0, \sum_{i=1}^k D_i\right\}$ satisfies $\tau \neq 0$. Let $\boldsymbol{\phi}_{\mathcal{B}_i}$ be the stationary distributions of $\mathcal{B}_i$, for $i = 1, ..k$. We claim that $\boldsymbol{\phi}_{\mathcal{G}} = \boldsymbol{\phi}_{\mathcal{B}_i}$, for $i = 1, ..., k$. This follows from the fact that $\mathcal{G}$ and $\mathcal{B}_i$ possesses the same infinitesimal generator Q (see (1.12)) and therefore the same stationary probability vectors π. It is immediate to check that expression (1.14) is the same for $\mathcal{G}$ and $\mathcal{B}_i$, since $D = \sum_{l=1}^k D_l$ coincides under the k considered models.

As it was proven in Section 4.2, for any $\kappa^i \in (0, \kappa_{\mathcal{B}_i})$, where $\kappa_{\mathcal{B}_i}$ is computed according to Proposition 4.2, there exists a unique $BMAP_2(2)$, $\widetilde{\mathcal{B}}_i(\kappa_{\mathcal{B}_i})$ equivalent to $\mathcal{B}_i$. This result leads to the next Theorem, where a $BMAP_2(k)$ for $k \geq 3$, is constructed from a specific choice of κ^i, $i = 1, ..., k$.

Theorem 4.1. *Consider a $BMAP_2(k)$, $k \geq 3$, defined by $\mathcal{G} = \{D_0, D_1, ..., D_k\}$ and the $BMAP_2(2)$s, $\mathcal{B}_1$,..., $\mathcal{B}_k$, as in (4.25). Let $\widetilde{\mathcal{B}}_1(\kappa)$,...,$\widetilde{\mathcal{B}}_k(\kappa)$ the equivalent $BMAP_2(2)$s to $\mathcal{B}_1$,..., $\mathcal{B}_k$ represented by*

$$\begin{aligned}
\widetilde{\mathcal{B}}_1(\kappa) &= \left\{\widetilde{D}_0(\kappa), \widetilde{D}_1(\kappa), \left(\widetilde{\sum_{i=2}^k D_i}\right)(\kappa)\right\}, \\
&\vdots \\
\widetilde{\mathcal{B}}_k(\kappa) &= \left\{\widetilde{D}_0(\kappa), \left(\widetilde{\sum_{i=1, i\neq k}^k D_i}\right)(\kappa), \widetilde{D}_k(\kappa)\right\},
\end{aligned}$$

where $0 < \kappa < min(\kappa_{\mathcal{B}_1}, ..., \kappa_{\mathcal{B}_k})$, being $\kappa_{\mathcal{B}_i}$ obtained from (4.23) or (4.24), for $i = 1, ..., k$. Then, the representation

$$\widetilde{\mathcal{G}}(\kappa) = \left\{\widetilde{D}_0(\kappa), \widetilde{D}_1(\kappa), ..., \widetilde{D}_k(\kappa)\right\}, \tag{4.26}$$

defines a $BMAP_2(k)$, for $k \geq 3$.

Proof. Note that the matrices in $\left\{\widetilde{D}_0(\kappa), \widetilde{D}_1(\kappa), ..., \widetilde{D}_k(\kappa)\right\}$ are rate matrices since they define the equivalent $BMAP_2(2)$s, $\widetilde{\mathcal{B}}_1(\kappa)$, ..., $\widetilde{\mathcal{B}}_k(\kappa)$. Therefore, for $\widetilde{\mathcal{G}}(\kappa)$ as in (4.26) to define a $BMAP_2(k)$, for $k \geq 3$, we have only to prove that $\left(\sum_{i=0}^k \widetilde{D}_i(\kappa)\right) \mathbf{e} = \mathbf{0}$. This would conclude the proof.

Consider the joint density function of the pair $(\boldsymbol{T}, \boldsymbol{B})$ in a general $BMAP_2(k)$ (see Klemm et al. [51]),

$$f_{\{D_0,...,D_k\}}(\mathbf{t}, \boldsymbol{b}) = \boldsymbol{\phi} e^{D_0 t_1} D_{b_1} \ldots e^{D_0 t_n} D_{b_n} \mathbf{e}, \tag{4.27}$$

where $\mathbf{t} = (t_1, \ldots, t_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ with $b_i \leq k$, for $i = 1, \ldots, n$, denote the sequence of inter-event times and batch sizes, respectively. Note that, $f_{\{D_0,\ldots,D_k\}}(\mathbf{t}, \boldsymbol{b})$ is the inverse of the Laplace Stieltjes transform given by (4.2). Since $\widetilde{\mathcal{B}}_i(\kappa)$ is equivalent to $\mathcal{B}_i$, for $i = 1, ..., k$, one has

$$\begin{aligned} f_{\left\{D_0,D_1,\sum_{i=2}^k D_i\right\}}(\mathbf{t},\boldsymbol{b}) &= f_{\left\{\widetilde{D}_0(\kappa),\widetilde{D}_1(\kappa),\widetilde{\left(\sum_{i=2}^k D_i\right)}(\kappa)\right\}}(\mathbf{t},\boldsymbol{b}), \\ &\vdots \\ f_{\left\{D_0,\sum_{i=1,i\neq k}^k D_i,D_k\right\}}(\mathbf{t},\boldsymbol{b}) &= f_{\left\{\widetilde{D}_0(\kappa),\widetilde{\left(\sum_{i=1,i\neq k}^k D_i\right)}(\kappa),\widetilde{D}_k(\kappa)\right\}}(\mathbf{t},\boldsymbol{b}), \end{aligned} \tag{4.28}$$

for all $\mathbf{t}$ and $\boldsymbol{b}$. Consider now the first equality in (4.28) concerning the equivalence between $\mathcal{B}_1$ and $\widetilde{\mathcal{B}}_1$, and take $\mathbf{t} = t$ and $\boldsymbol{b} = 2$. Then,

$$\begin{aligned} \boldsymbol{\phi}_{\widetilde{\mathcal{B}}_1} e^{\widetilde{D}_0 t} \widetilde{\left(\sum_{i=2}^k D_i\right)}(\kappa)\mathbf{e} &= f_{\left\{\widetilde{D}_0(\kappa),\widetilde{D}_1(\kappa),\widetilde{\left(\sum_{i=2}^k D_i\right)}(\kappa)\right\}}(\mathbf{t},\boldsymbol{b}) \\ &= f_{\left\{D_0,D_1,\sum_{i=2}^k D_i\right\}}(\mathbf{t},\boldsymbol{b}), \\ &= \boldsymbol{\phi}_{\mathcal{B}_1} e^{D_0 t} \left(\sum_{i=2}^k D_i\right)\mathbf{e} \\ &= \sum_{i=2}^k \boldsymbol{\phi}_{\mathcal{B}_1} e^{D_0 t} D_i \mathbf{e} \\ &= \sum_{i=2}^k \boldsymbol{\phi}_{\widetilde{\mathcal{B}}_1} e^{\widetilde{D}_0 t} \widetilde{D}_i(\kappa)\mathbf{e} \qquad (4.29) \\ &= \boldsymbol{\phi}_{\widetilde{\mathcal{B}}_1} e^{\widetilde{D}_0 t} \left(\sum_{i=2}^k \widetilde{D}_i(\kappa)\right)\mathbf{e}, \end{aligned}$$

where (4.29) follows from $\mathcal{B}_i \equiv \widetilde{\mathcal{B}}_i$, for $i = 2, ..., k$. Since given the pair $(\mathcal{B}_1, \kappa)$ there exists a unique equivalent process, $\widetilde{\mathcal{B}}_1(\kappa)$, then necessarily

$$\widetilde{\left(\sum_{i=2}^k D_i\right)}(\kappa) = \sum_{i=2}^k \widetilde{D}_i(\kappa). \tag{4.30}$$

Since $\widetilde{\mathcal{B}}_1(\kappa)$ is a $BMAP_2(2)$, then

$$\left(\widetilde{D}_0(\kappa) + \widetilde{D}_1(\kappa) + \left(\sum_{i=2}^k \widetilde{D}_i\right)(\kappa)\right)\mathbf{e} = \mathbf{0},$$

which combined with (4.30) finally leads to

$$\left(\sum_{i=0}^{k} \widetilde{D}_i(\kappa)\right)\mathbf{e} = \mathbf{0}.$$

□

The next theorem proves the non-identifiability of $BMAP_2(k)$, for $k \geq 3$, through the decomposition of $\mathcal{G}$ into k $\mathcal{B}_i$s.

Theorem 4.2. *The stationary $BMAP_2(k)$ is a non-identifiable process for $k \geq 3$.*

Proof. The proof relies on the fact that for a given $BMAP_2(k)$, $\mathcal{G} = \{D_0, D_1, \ldots, D_k\}$, there exists a $BMAP_2(k)$, $\widetilde{\mathcal{G}} = \widetilde{\mathcal{G}}(\kappa)$ as in (4.26) such that $\mathcal{G}$ is equivalent to $\widetilde{\mathcal{G}}$.

We proceed now to show that if $0 < \kappa < \min(\kappa_{\mathcal{B}_1}, ..., \kappa_{\mathcal{B}_k})$ and $\widetilde{\mathcal{G}}$ constructed as in (4.26), then $\widetilde{\mathcal{G}}$ is equivalent to $\mathcal{G}$. Or, alternatively, the equality of Laplace transforms (4.5) holds,

$$f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s}, \boldsymbol{z})_{\mathcal{G}} = f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(\mathbf{s}, \boldsymbol{z})_{\widetilde{\mathcal{G}}}.$$

For simplicity, we prove the equality of Laplace transforms (4.5) for $k = 3$. The generalization for $k \geq 4$ is straightforward. Define $\Delta(s_i) = (s_i I - D_0)^{-1}$ and $\xi(z_i) = \sum_{l=1}^{k} D_l z_i^l$, for $i = 1, ..., n$. Consider $n = 1$, then

$$\begin{aligned}
f^*_{\boldsymbol{T},\boldsymbol{B}}(s_1, z_1)_{\mathcal{G}} &= \boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)\xi(z_1)\mathbf{e} \\
&= \boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)\left(D_1 z_1 + D_2 {z_1}^2 + D_3 {z_1}^3\right)\mathbf{e} \\
&= \boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)\left(D_1 z_1 + (D_2 + D_3){z_1}^2\right)\mathbf{e} \\
&\quad + {z_1}^2\boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)\left((D_1 + D_2)z_1 + D_3 {z_1}^2\right)\mathbf{e} \\
&\quad - {z_1}^2\boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)(D_1 + D_2 + D_3)\mathbf{e} \\
&= f^*_{\boldsymbol{T},\boldsymbol{B}}(s_1, z_1)_{\mathcal{B}_1} + z_1 f^*_{\boldsymbol{T},\boldsymbol{B}}(s_1, z_1)_{\mathcal{B}_3} - {z_1}^2 f^*_{\boldsymbol{T},\boldsymbol{B}}(s_1, 1)_{\mathcal{B}_2} \\
&= f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(s_1, z_1)_{\widetilde{\mathcal{B}}_1} + z_1 f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(s_1, z_1)_{\widetilde{\mathcal{B}}_3} - {z_1}^2 f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(s_1, 1)_{\widetilde{\mathcal{B}}_2} \\
&= \boldsymbol{\phi}_{\widetilde{\mathcal{G}}}\widetilde{\Delta}(s_1)\widetilde{\xi}(z_1)\mathbf{e} \\
&= f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(s_1, z_1)_{\widetilde{\mathcal{G}}},
\end{aligned}$$

where the equivalence between $\mathcal{B}_i$ and $\widetilde{\mathcal{B}}_i$ for $i = 1, 2, 3$, and the equality of the stationary distributions, $\boldsymbol{\phi}_{\mathcal{G}} = \boldsymbol{\phi}_{\mathcal{B}_i}$ and $\boldsymbol{\phi}_{\widetilde{\mathcal{G}}} = \boldsymbol{\phi}_{\widetilde{\mathcal{B}}_i}$, for $i = 1, 2, 3$, have been applied. We now proceed by induction on n,

$$\begin{aligned}
f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s},\boldsymbol{z})_{\mathcal{G}} &= \boldsymbol{\phi}_{\mathcal{G}}\Delta(s_1)\xi(z_1)\ldots\Delta(s_n)\xi(z_n)\mathbf{e}\\
&= \boldsymbol{\phi}_{\mathcal{G}}\chi(n-1)\Delta(s_n)\left(D_1 z_n + D_2 {z_n}^2 + D_3 {z_n}^3\right)\mathbf{e}\\
&= \boldsymbol{\phi}_{\mathcal{G}}\chi(n-1)\Delta(s_n)\left(D_1 z_n + (D_2+D_3){z_n}^2\right)\mathbf{e}\\
&\quad + {z_n}^2\boldsymbol{\phi}_{\mathcal{G}}\chi(n-1)\Delta(s_n)\left((D_1+D_2)z_n + D_3{z_n}^2\right)\mathbf{e}\\
&\quad - {z_n}^2\boldsymbol{\phi}_{\mathcal{G}}\chi(n-1)\Delta(s_n)(D_1+D_2+D_3)\mathbf{e}\\
&= f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s},\boldsymbol{z})_{\mathcal{B}_1} + z_n f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s},\boldsymbol{z})_{\mathcal{B}_3} - {z_n}^2 f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s},1)_{\mathcal{B}_2}\\
&= f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(\mathbf{s},\boldsymbol{z})_{\widetilde{\mathcal{B}}_1} + z_n f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(\mathbf{s},\boldsymbol{z})_{\widetilde{\mathcal{B}}_3} - {z_n}^2 f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(\mathbf{s},1)_{\widetilde{\mathcal{B}}_2}\\
&= \boldsymbol{\phi}_{\widetilde{\mathcal{G}}}\widetilde{\chi}(n-1)\widetilde{\Delta}(s_n)\widetilde{\xi}(z_n)\mathbf{e}\\
&= f^*_{\widetilde{\boldsymbol{T}},\widetilde{\boldsymbol{B}}}(\mathbf{s},\boldsymbol{z})_{\widetilde{\mathcal{G}}},
\end{aligned}$$

where $\chi(n-1) = \prod_{i=1}^{n-1}\Delta(s_i)\xi(z_i)$. In consequence, $\widetilde{\mathcal{G}}$ and $\mathcal{G}$ are equivalent. □

In the next example we illustrate our approach for finding equivalent $BMAP_2(3)$ to a given fixed one.

Example 4.2. *Consider the $BMAP_2(3)$ defined by*

$$\mathcal{G}: \quad D_0 = \begin{pmatrix} -5.5 & 0.5 \\ 0.55 & -4 \end{pmatrix},\ D_1 = \begin{pmatrix} 0.85 & 0.65 \\ 0.05 & 0.9 \end{pmatrix},\ D_2 = \begin{pmatrix} 1.2 & 0.55 \\ 0.65 & 0.45 \end{pmatrix},\ D_3 = \begin{pmatrix} 1.4 & 0.35 \\ 0.85 & 0.55 \end{pmatrix},$$

whose underlying $BMAP_2(2)$s are

$$\mathcal{B}_1: \qquad D_0 = \begin{pmatrix} -5.5 & 0.5 \\ 0.55 & -4 \end{pmatrix},\ D_1 = \begin{pmatrix} 0.85 & 0.65 \\ 0.05 & 0.9 \end{pmatrix},\ D_2 + D_3 = \begin{pmatrix} 2.6 & 0.9 \\ 1.5 & 1 \end{pmatrix},$$

$$\mathcal{B}_2: \qquad D_0 = \begin{pmatrix} -5.5 & 0.5 \\ 0.55 & -4 \end{pmatrix},\ D_1 + D_3 = \begin{pmatrix} 2.25 & 1 \\ 0.9 & 1.45 \end{pmatrix},\ D_2 = \begin{pmatrix} 1.2 & 0.55 \\ 0.65 & 0.45 \end{pmatrix},$$

$$\mathcal{B}_3: \qquad D_0 = \begin{pmatrix} -5.5 & 0.5 \\ 0.55 & -4 \end{pmatrix},\ D_1 + D_2 = \begin{pmatrix} 2.05 & 1.2 \\ 0.7 & 1.35 \end{pmatrix},\ D_3 = \begin{pmatrix} 1.4 & 0.35 \\ 0.85 & 0.55 \end{pmatrix},$$

where $\{\kappa_{\mathcal{B}_1}, \kappa_{\mathcal{B}_2}, \kappa_{\mathcal{B}_3}\} = \{0.2239, 0.2799, 0.1724\}$ are the respective upper bounds obtained from Proposition 4.2.

Let $\kappa = 0.05$ that verifies

$$0 < \kappa < min(0.2239, 0.2799, 0.1724),$$

we can construct the equivalent $\widetilde{B}_i s$, for $i = 1, 2, 3$,

$$\widetilde{\mathcal{B}}_1 : \quad \widetilde{D}_0(\kappa) = \begin{pmatrix} -5.45 & 0.5792 \\ 0.6 & -4.05 \end{pmatrix}, \qquad \widetilde{D}_1(\kappa) = \begin{pmatrix} 0.8545 & 0.5996 \\ 0.0545 & 0.8955 \end{pmatrix},$$

$$(\widetilde{D_2 + D_3})(\kappa) = \begin{pmatrix} 2.7364 & 0.6803 \\ 1.6364 & 0.8636 \end{pmatrix},$$

$$\widetilde{\mathcal{B}}_2 : \quad \widetilde{D}_0(\kappa) = \begin{pmatrix} -5.45 & 0.5792 \\ 0.6 & -4.05 \end{pmatrix}, \qquad \widetilde{D}_2(\kappa) = \begin{pmatrix} 1.2591 & 0.4367 \\ 0.7091 & 0.3909 \end{pmatrix},$$

$$(\widetilde{D_1 + D_3})(\kappa) = \begin{pmatrix} 2.3318 & 0.8432 \\ 0.9818 & 1.3682 \end{pmatrix},$$

$$\widetilde{\mathcal{B}}_3 : \quad \widetilde{D}_0(\kappa) = \begin{pmatrix} -5.45 & 0.5792 \\ 0.6 & -4.05 \end{pmatrix}, \qquad \widetilde{D}_3(\kappa) = \begin{pmatrix} 1.4773 & 0.2436 \\ 0.9273 & 0.4727 \end{pmatrix},$$

$$(\widetilde{D_1 + D_2})(\kappa) = \begin{pmatrix} 2.1136 & 1.0364 \\ 0.7636 & 1.2864 \end{pmatrix}.$$

Then, from Theorem 4.1 an equivalent BMAP$_2$(3), $\widetilde{\mathcal{G}}(\kappa)$, can be constructed as follows,

$$\widetilde{\mathcal{G}} : \quad \widetilde{D}_0(\kappa) = \begin{pmatrix} -5.45 & 0.5792 \\ 0.6 & -4.05 \end{pmatrix}, \widetilde{D}_1(\kappa) = \begin{pmatrix} 0.8545 & 0.5996 \\ 0.0545 & 0.8955 \end{pmatrix}, \widetilde{D}_2(\kappa) = \begin{pmatrix} 1.2591 & 0.4367 \\ 0.7091 & 0.3909 \end{pmatrix},$$

$$\widetilde{D}_3 = \begin{pmatrix} 1.4773 & 0.2436 \\ 0.9273 & 0.4727 \end{pmatrix},$$

where $\widetilde{D}_0(\kappa)$ is obtained from the equivalence of the underlying MAP$_2$ and $\widetilde{D}_i(\kappa)$ is obtained from $\widetilde{\mathcal{B}}_i$ for $i = 1, 2, 3$.

4.4 Chapter summary

This chapter deepens the understanding of the identifiability of the *BMAP* process, a relevant aspect not only from the theoretical viewpoint, but also when inference for the process is to be undertaken. Specifically, it proves that the two-state stationary *BMAP* or $BMAP_2(2)$ is non-identifiable, extending previous works focused on the case $k = 1$. The main result of this chapter is the introduction of a method that shows how to build an equivalent stationary $BMAP_2(k)$s, to any given fixed one, which is derived from equality of the Laplace transforms for $n = 1, 2$, and for $k \geq 3$ this method is based on the construction of equivalent $BMAP_2(k)$s, and on the decomposition of a $BMAP_2(k)$ into k $BMAP_2(2)$s.

In this chapter, similarly as in previous works considering analogous problems, the concept of identifiability is expressed in terms of the equality in distribution of a reduced number of components of the process (inter-event times and batch sizes). Since the sequences of inter-event times and batch sizes constitute the only available information for estimation purposes, other elements of the process which commonly remain unobserved as the times between transitions without event occurrences, or the sequence of visited states, are not taken into account in the identifiability definition. This fact in combination with the high number of parameters defining the processes induce the lack of identifiability of *BMAP*s.

Appendices

4.A Proof of Lemma 4.1

Let $\boldsymbol{T} = (T_1, \ldots, T_n)$, $\boldsymbol{B} = (B_1, \ldots, B_n)$, $\mathbf{s} = (s_1, \ldots, s_n)$, $\boldsymbol{z} = (z_1, \ldots, z_n)$, $\mathbf{t} = (t_1, \ldots, t_n)$, $\boldsymbol{b} = (b_1, \ldots, b_n)$. Then, expression (4.2) easily follows from the definition of the moment generating function of $(\boldsymbol{T}, \boldsymbol{B})$,

$$f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s}, \boldsymbol{z}) = \int_0^\infty e^{-s_1 t_1} \ldots e^{-s_n t_n} \prod_{j=1}^{n} \left(\sum_{b_j=1}^{\infty} z_j^{b_j} \right) f_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{t}, \boldsymbol{b})\, d\mathbf{t}, \tag{4.31}$$

We have that the cumulative distribution function for (T_n, B_n) is

$$F_{T_n,B_n}(t_n, b_n) = \boldsymbol{\phi}(I - e^{D_0 t_n})(-D_0)^{-1} D_{b_n} \mathbf{e}, \quad \text{for } t_n, b_n \geq 0.$$

Therefore, the joint cumulative distribution function is,

$$F_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{t}, \boldsymbol{b}) = \boldsymbol{\phi}(I - e^{D_0 t_1})(-D_0)^{-1} D_{b_1} \ldots (I - e^{D_0 t_n})(-D_0)^{-1} D_{b_n} \mathbf{e}.$$

and the joint density function of $(\boldsymbol{T},\boldsymbol{B})$, $f_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{t},\boldsymbol{b})$, is given by (4.27).

Now, after substituting $f_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{t},\boldsymbol{b})$ in (4.31) and regrouping, we obtain

$$\begin{aligned} f^*_{\boldsymbol{T},\boldsymbol{B}}(\mathbf{s},\boldsymbol{z}) &= \boldsymbol{\phi}\int_0^\infty e^{-s_1t_1}e^{D_0t_1}\left(\sum_{b_1=1}^\infty D_{b_1}z_1^{b_1}\right)\dots e^{-s_nt_n}e^{D_0t_n}\left(\sum_{b_n=1}^\infty D_{b_n}z_n^{b_n}\right)\mathbf{e}d\mathbf{t} \\ &= \boldsymbol{\phi}(s_1I-D_0)^{-1}\left(\sum_{b_1=1}^\infty D_{b_1}z_1^{b_1}\right)\dots(s_nI-D_0)^{-1}\left(\sum_{b_n=1}^\infty D_{b_n}z_n^{b_n}\right)\mathbf{e}. \end{aligned}$$

□

4.B Proof of Lemma 4.2

Let $\boldsymbol{T}=(T_1,\dots,T_n)$, $\boldsymbol{B}=(B_1,\dots,B_n)$, $\mathbf{s}=(s_1,\dots,s_n)$ and $\boldsymbol{z}=(z_1,\dots,z_n)$. Since

$$E[B_1B_n]=\frac{\partial f^*_{(\boldsymbol{T},\boldsymbol{B})}(\mathbf{s},\boldsymbol{z})}{\partial z_1\partial z_n}\bigg|_{\mathbf{s}=\mathbf{0};\boldsymbol{z}=\mathbf{1}},$$

then, it follows from (4.2) that

$$\begin{aligned} E[B_1B_n] &= \boldsymbol{\phi}(-D_0)^{-1}\left(\sum_{l=1}^k lD_lz_1^{l-1}\right)\kappa(-D_0)^{-1}\left(\sum_{l=1}^k lD_lz_n^{l-1}\right)\mathbf{e}\bigg|_{\mathbf{s}=\mathbf{0};\boldsymbol{z}=\mathbf{1}} \\ &= \boldsymbol{\phi}(-D_0)^{-1}D_1^\star[(-D_0)^{-1}D]^{n-2}(-D_0)^{-1}D_1^\star\mathbf{e}, \end{aligned}$$

where $\kappa=\sum_{m=2}^{n-1}(s_mI-D_0)^{-1}\left(\sum_{l=1}^k D_lz_m^l\right)$. Therefore (4.3) is obtained.

□

4.C Proof of Formula (4.17)

It can be easily seen that, after calculation, condition (4.12) holds if and only if

$$\begin{aligned}
\gamma &= \widetilde{\gamma}, \\
\eta - \beta + \gamma\widetilde{\upsilon} &= \widetilde{\eta} - \widetilde{\beta} + \widetilde{\gamma}\upsilon, \\
\widetilde{\eta}\gamma - \beta\widetilde{\upsilon} + \eta\widetilde{\upsilon} &= \eta\widetilde{\gamma} - \widetilde{\beta}\upsilon + \widetilde{\eta}\upsilon, \\
\eta\widetilde{\eta} - \beta\widetilde{\eta} &= \widetilde{\eta}\eta - \widetilde{\beta}\eta, \\
\alpha &= \widetilde{\alpha}, \\
\beta + \alpha\widetilde{\upsilon} &= \widetilde{\beta} + \widetilde{\alpha}\upsilon, \\
\alpha\widetilde{\eta} + \beta\widetilde{\upsilon} &= \widetilde{\alpha}\eta + \widetilde{\beta}\upsilon, \\
\beta\widetilde{\eta} &= \widetilde{\beta}\eta,
\end{aligned} \tag{4.32}$$

the system (4.32) holds.

Considering that $\alpha = \widetilde{\alpha}$ and $\gamma = \widetilde{\gamma}$, we solve the linear system (4.32) with unknowns, $\widetilde{\beta}$, $\widetilde{\upsilon}$ and $\widetilde{\eta}$. We used the Gaussian elimination method to solve this problem, and obtained

$$\begin{aligned}
\widetilde{\beta} - \alpha\widetilde{\upsilon} &= \beta - \alpha\upsilon, \\
\widetilde{\upsilon}(\alpha\eta - \beta(\alpha + \gamma)) &= \upsilon(\alpha\eta - \beta(\alpha + \gamma)), \\
\widetilde{\eta} - \widetilde{\upsilon}(\alpha + \gamma) &= \eta - \upsilon(\alpha + \gamma), \\
\widetilde{\upsilon}(\beta + \alpha(\alpha + \gamma - \upsilon)) &= \upsilon(\beta + \alpha(\alpha + \gamma - \upsilon)), \\
\widetilde{\upsilon}(\eta + (\gamma - \upsilon + \alpha)(\alpha + \gamma)) &= \upsilon(\eta + (\gamma - \upsilon + \alpha)(\alpha + \gamma)).
\end{aligned} \tag{4.33}$$

Recall that

$$\begin{aligned}
C_1 &= \alpha\eta - \beta(\alpha + \gamma), \\
C_2 &= \beta + \alpha(\alpha + \gamma - \upsilon), \\
C_3 &= \eta - \beta + \gamma(\gamma - \upsilon + \alpha),
\end{aligned}$$

note that if either $C1 \neq 0$ or $C2 \neq 0$ or $C3 \neq 0$ then (4.33) implies

$$\alpha = \widetilde{\alpha}, \qquad \beta = \widetilde{\beta}, \qquad \gamma = \widetilde{\gamma}, \qquad \eta = \widetilde{\eta}, \qquad \upsilon = \widetilde{\upsilon}.$$

Consider now the condition (4.15) which holds if and only if,

$$\frac{z_1z_2(s_1\delta_1+s_2\delta_2+s_1s_2\delta_3+\delta_4)+z_1z_2^2(s_1(\alpha\eta-\delta_1)+s_2\delta_5+s_1s_2\delta_6)}{(s_1^2+s_2^2)\eta+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\upsilon+(s_1+s_2)\eta\upsilon+s_1s_2\upsilon^2+\eta^2}+$$
$$\frac{z_1z_2^2(\beta\eta-\delta_4)+z_1^2z_2(s_1\delta_7+s_2\delta_8+s_1s_2\delta_9+\delta_{10})}{(s_1^2+s_2^2)\eta+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\upsilon+(s_1+s_2)\eta\upsilon+s_1s_2\upsilon^2+\eta^2}+$$
$$\frac{z_1^2z_2^2(s_1(\eta\gamma-\delta_7)+s_2\delta_{11}+s_1s_2\delta_{12}+\eta^2-\eta\beta-\delta_{10})}{(s_1^2+s_2^2)\eta+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\upsilon+(s_1+s_2)\eta\upsilon+s_1s_2\upsilon^2+\eta^2}$$
$$=$$
$$\frac{z_1z_2(s_1\widetilde{\delta}_1+s_2\widetilde{\delta}_2+s_1s_2\widetilde{\delta}_3+\widetilde{\delta}_4)+z_1z_2^2(s_1(\widetilde{\alpha}\widetilde{\eta}-\widetilde{\delta}_1)+s_2\widetilde{\delta}_5+s_1s_2\widetilde{\delta}_6)}{(s_1^2+s_2^2)\widetilde{\eta}+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\widetilde{\upsilon}+(s_1+s_2)\widetilde{\eta}\widetilde{\upsilon}+s_1s_2\widetilde{\upsilon}^2+\widetilde{\eta}^2}+$$
$$\frac{z_1z_2^2(\widetilde{\beta}\widetilde{\eta}-\widetilde{\delta}_4)+z_1^2z_2(s_1\widetilde{\delta}_7+s_2\widetilde{\delta}_8+s_1s_2\widetilde{\delta}_9+\widetilde{\delta}_{10})}{(s_1^2+s_2^2)\widetilde{\eta}+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\widetilde{\upsilon}+(s_1+s_2)\widetilde{\eta}\widetilde{\upsilon}+s_1s_2\widetilde{\upsilon}^2+\widetilde{\eta}^2}+$$
$$\frac{z_1^2z_2^2(s_1(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)+s_2\widetilde{\delta}_{11}+s_1s_2\widetilde{\delta}_{12}+\widetilde{\eta}^2-\widetilde{\eta}\widetilde{\beta}-\widetilde{\delta}_{10})}{(s_1^2+s_2^2)\widetilde{\eta}+s_1^2s_2^2+(s_1^2s_2+s_1s_2^2)\widetilde{\upsilon}+(s_1+s_2)\widetilde{\eta}\widetilde{\upsilon}+s_1s_2\widetilde{\upsilon}^2+\widetilde{\eta}^2},$$

which is equivalent that the following holds,

$$\begin{aligned}
\delta_{12} &= \widetilde{\delta_{12}}\\
\delta_{12}\widetilde{\upsilon}+(\eta\gamma-\delta_7) &= \widetilde{\delta_{12}}\upsilon+(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\\
\delta_{12}\widetilde{\eta}+(\eta\gamma-\delta_7)\widetilde{\upsilon} &= \widetilde{\delta_{12}}\eta+(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\upsilon\\
(\eta\gamma-\delta_7)\widetilde{\eta} &= (\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\eta\\
\delta_{11}+\delta_{12}\widetilde{\upsilon} &= \widetilde{\delta_{11}}+\widetilde{\delta_{12}}\upsilon\\
(\eta\gamma-\delta_7+\delta_{11})\widetilde{\upsilon}+\eta^2-\eta\beta-\delta_{10}+\delta_{12}\widetilde{\upsilon}^2 &= (\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7+\widetilde{\delta_{11}})\upsilon+\widetilde{\eta}^2-\widetilde{\eta}\widetilde{\beta}-\widetilde{\delta_{10}}+\widetilde{\delta_{12}}\upsilon^2\\
\delta_{11}\widetilde{\eta}+(\eta^2-\eta\beta-\delta_{10})\widetilde{\upsilon}+(\eta\gamma-\delta_7)\widetilde{\upsilon}^2+\delta_{12}\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta_{11}}\eta+(\widetilde{\eta}^2-\widetilde{\eta}\widetilde{\beta}-\widetilde{\delta_{10}})\upsilon+(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\upsilon^2+\widetilde{\delta_{12}}\eta\upsilon\\
(\eta(\eta-\beta)-\delta_{10})\widetilde{\eta}+(\eta\gamma-\delta_7)\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\eta}(\widetilde{\eta}-\widetilde{\beta})-\widetilde{\delta_{10}})\eta+(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\eta\upsilon\\
\delta_{12}\widetilde{\eta}+\delta_{11}\widetilde{\upsilon} &= \widetilde{\delta_{12}}\eta+\widetilde{\delta_{11}}\upsilon\\
(\eta\gamma-\delta_7)\widetilde{\eta}+(\eta^2-\eta\beta-\delta_{10})\widetilde{\upsilon}+\delta_{11}\widetilde{\upsilon}^2+\delta_{12}\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\eta+(\widetilde{\eta}^2-\widetilde{\eta}\widetilde{\beta}-\widetilde{\delta_{10}})\upsilon+\widetilde{\delta_{11}}\upsilon^2+\widetilde{\delta_{12}}\eta\upsilon\\
\delta_{12}\widetilde{\eta}^2+(\eta^2-\eta\beta-\delta_{10})\widetilde{\upsilon}^2+(\eta\gamma-\delta_7+\delta_{11})\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta_{12}}\eta^2+(\widetilde{\eta}^2-\widetilde{\eta}\widetilde{\beta}-\widetilde{\delta_{10}})\upsilon^2+(\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7+\widetilde{\delta_{11}})\eta\upsilon\\
(\eta\gamma-\delta_7)\widetilde{\eta}^2+(\eta(\eta-\beta)-\delta_{10})\widetilde{\upsilon}\widetilde{\eta} &= (\widetilde{\eta}\widetilde{\gamma}-\widetilde{\delta}_7)\eta^2+(\widetilde{\eta}(\widetilde{\eta}-\widetilde{\beta})-\widetilde{\delta_{10}})\upsilon\eta\\
\delta_{11}\widetilde{\eta} &= \widetilde{\delta_{11}}\eta\\
(\eta(\eta-\beta)-\delta_{10})\widetilde{\eta}+\delta_{11}\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\eta}(\widetilde{\eta}-\widetilde{\beta})-\widetilde{\delta_{10}})\eta+\widetilde{\delta_{11}}\eta\upsilon\\
\delta_{11}\widetilde{\eta}^2+(\eta(\eta-\beta)-\delta_{10})\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta_{11}}\eta^2+(\widetilde{\eta}(\widetilde{\eta}-\widetilde{\beta})-\widetilde{\delta_{10}})\upsilon\eta\\
(\eta(\eta-\beta)-\delta_{10})\widetilde{\eta}^2 &= (\widetilde{\eta}(\widetilde{\eta}-\widetilde{\beta})-\widetilde{\delta_{10}})\eta^2\\
\delta_9 &= \widetilde{\delta}_9
\end{aligned}$$

$$\begin{aligned}
\delta_7 + \delta_9\widetilde{\upsilon} &= \widetilde{\delta}_7 + \widetilde{\delta}_9\upsilon \\
\delta_9\widetilde{\eta} + \delta_7\widetilde{\upsilon} &= \widetilde{\delta}_9\eta + \widetilde{\delta}_7\upsilon \\
\delta_7\widetilde{\eta} &= \widetilde{\delta}_7\eta \\
\delta_8 + \delta_9\widetilde{\upsilon} &= \widetilde{\delta}_8 + \widetilde{\delta}_9\upsilon \\
\delta_{10} + (\delta_7 + \delta_8)\widetilde{\upsilon} + \delta_9\widetilde{\upsilon}^2 &= \widetilde{\delta_{10}} + (\widetilde{\delta}_7 + \widetilde{\delta}_8)\upsilon + \widetilde{\delta}_9\upsilon^2 \\
\delta_8\widetilde{\eta} + \delta_{10}\widetilde{\upsilon} + \delta_7\widetilde{\upsilon}^2 + \delta_9\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_8\eta + \widetilde{\delta_{10}}\upsilon + \widetilde{\delta}_7\upsilon^2 + \widetilde{\delta}_9\eta\upsilon \\
\delta_{10}\widetilde{\eta} + \delta_7\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta_{10}}\eta + \widetilde{\delta}_7\eta\upsilon \\
\delta_9\widetilde{\eta} + \delta_8\widetilde{\upsilon} &= \widetilde{\delta}_9\eta + \widetilde{\delta}_8\upsilon \\
\delta_7\widetilde{\eta} + \delta_{10}\widetilde{\upsilon} + \delta_8\widetilde{\upsilon}^2 + \delta_9\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_7\eta + \widetilde{\delta_{10}}\upsilon + \widetilde{\delta}_8\upsilon^2 + \widetilde{\delta}_9\eta\upsilon \\
\delta_9\widetilde{\eta}^2 + \delta_{10}\widetilde{\upsilon}^2 + (\delta_7 + \delta_8)\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_9\eta^2 + \widetilde{\delta_{10}}\upsilon^2 + (\widetilde{\delta}_7 + \widetilde{\delta}_8)\eta\upsilon \\
\delta_7\widetilde{\eta}^2 + \delta_{10}\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta}_7\eta^2 + \widetilde{\delta_{10}}\upsilon\eta \\
\delta_8\widetilde{\eta} &= \widetilde{\delta}_8\eta \\
\delta_{10}\widetilde{\eta} + \delta_8\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta_{10}}\eta + \widetilde{\delta}_8\eta\upsilon \\
\delta_8\widetilde{\eta}^2 + \delta_{10}\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta}_8\eta^2 + \widetilde{\delta_{10}}\upsilon\eta \\
\delta_{10}\widetilde{\eta}^2 &= \widetilde{\delta_{10}}\eta^2 \\
\delta_6 &= \widetilde{\delta}_6 \\
(\alpha\eta - \delta_1) + \delta_6\widetilde{\upsilon} &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1) + \widetilde{\delta}_6\upsilon \\
\delta_6\widetilde{\eta} + (\alpha\eta - \delta_1)\widetilde{\upsilon} &= \widetilde{\delta}_6\eta + (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\upsilon \\
(\alpha\eta - \delta_1)\widetilde{\eta} &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\eta \\
\delta_5 + \delta_6\widetilde{\upsilon} &= \widetilde{\delta}_5 + \widetilde{\delta}_6\upsilon \\
(\alpha\eta - \delta_1 + \delta_5)\widetilde{\upsilon} + \beta\eta - \delta_4 + \delta_6\widetilde{\upsilon}^2 &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1 + \widetilde{\delta}_5)\upsilon + \widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4 + \widetilde{\delta}_6\upsilon^2 \\
(\beta\eta - \delta_4)\widetilde{\upsilon} + (\alpha\eta - \delta_1)\widetilde{\upsilon}^2 + \delta_5\widetilde{\eta} + \delta_6\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\upsilon + (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\upsilon^2 + \widetilde{\delta}_5\eta + \widetilde{\delta}_6\eta\upsilon \\
(\beta\eta - \delta_4)\widetilde{\eta} + (\alpha\eta - \delta_1)\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\eta + (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\eta\upsilon \\
\delta_6\widetilde{\eta} + \delta_5\widetilde{\upsilon} &= \widetilde{\delta}_6\eta + \widetilde{\delta}_5\upsilon \\
(\alpha\eta - \delta_1)\widetilde{\eta} + (\beta\eta - \delta_4)\widetilde{\upsilon} + \delta_5\widetilde{\upsilon}^2 + \delta_6\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\eta + (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\upsilon + \widetilde{\delta}_5\upsilon^2 + \widetilde{\delta}_6\eta\upsilon \\
(\alpha\eta - \delta_1 + \delta_5)\widetilde{\eta}\widetilde{\upsilon} + \delta_6\widetilde{\eta}^2 + (\beta\eta - \delta_4)\widetilde{\upsilon}^2 &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1 + \widetilde{\delta}_5)\eta\upsilon + \widetilde{\delta}_6\eta^2 + (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\upsilon^2 \\
(\alpha\eta - \delta_1)\widetilde{\eta}^2 + (\beta\eta - \delta_4)\widetilde{\upsilon}\widetilde{\eta} &= (\widetilde{\alpha}\widetilde{\eta} - \widetilde{\delta}_1)\eta^2 + (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\upsilon\eta \\
\delta_5\widetilde{\eta} &= \widetilde{\delta}_5\eta \\
(\beta\eta - \delta_4)\widetilde{\eta} + \delta_5\widetilde{\eta}\widetilde{\upsilon} &= (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\eta + \widetilde{\delta}_5\eta\upsilon \\
\delta_5\widetilde{\eta}^2 + (\beta\eta - \delta_4)\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta}_5\eta^2 + (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\upsilon\eta
\end{aligned}$$

$$
\begin{aligned}
(\beta\eta - \delta_4)\widetilde{\eta}^2 &= (\widetilde{\beta}\widetilde{\eta} - \widetilde{\delta}_4)\eta^2 \\
\delta_3 &= \widetilde{\delta}_3 \\
\delta_1 + \delta_3\widetilde{\upsilon} &= \widetilde{\delta}_1 + \widetilde{\delta}_3\upsilon \\
\delta_3\widetilde{\eta} + \delta_1\widetilde{\upsilon} &= \widetilde{\delta}_3\eta + \widetilde{\delta}_1\upsilon \\
\delta_1\widetilde{\eta} &= \widetilde{\delta}_1\eta \\
\delta_2 + \delta_3\widetilde{\upsilon} &= \widetilde{\delta}_2 + \widetilde{\delta}_3\upsilon \\
\delta_4 + (\delta_1 + \delta_2)\widetilde{\upsilon} + \delta_3\widetilde{\upsilon}^2 &= \widetilde{\delta}_4 + (\widetilde{\delta}_1 + \widetilde{\delta}_2)\upsilon + \widetilde{\delta}_3\upsilon^2 \\
\delta_2\widetilde{\eta} + \delta_4\widetilde{\upsilon} + \delta_1\widetilde{\upsilon}^2 + \delta_3\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_2\eta + \widetilde{\delta}_4\upsilon + \widetilde{\delta}_1\upsilon^2 + \widetilde{\delta}_3\eta\upsilon \\
\delta_4\widetilde{\eta} + \delta_1\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_4\eta + \widetilde{\delta}_1\eta\upsilon \\
\delta_3\widetilde{\eta} + \delta_2\widetilde{\upsilon} &= \widetilde{\delta}_3\eta + \widetilde{\delta}_2\upsilon \\
\delta_1\widetilde{\eta} + \delta_4\widetilde{\upsilon} + \delta_2\widetilde{\upsilon}^2 + \delta_3\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_1\eta + \widetilde{\delta}_4\upsilon + \widetilde{\delta}_2\upsilon^2 + \widetilde{\delta}_3\eta\upsilon \\
\delta_3\widetilde{\eta}^2 + \delta_4\widetilde{\upsilon}^2 + (\delta_1 + \delta_2)\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_3\eta^2 + \widetilde{\delta}_4\upsilon^2 + (\widetilde{\delta}_1 + \widetilde{\delta}_2)\eta\upsilon \\
\delta_1\widetilde{\eta}^2 + \delta_4\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta}_1\eta^2 + \widetilde{\delta}_4\upsilon\eta \\
\delta_2\widetilde{\eta} &= \widetilde{\delta}_2\eta \\
\delta_4\widetilde{\eta} + \delta_2\widetilde{\eta}\widetilde{\upsilon} &= \widetilde{\delta}_4\eta + \widetilde{\delta}_2\eta\upsilon \\
\delta_2\widetilde{\eta}^2 + \delta_4\widetilde{\upsilon}\widetilde{\eta} &= \widetilde{\delta}_2\eta^2 + \widetilde{\delta}_4\upsilon\eta \\
\delta_4\widetilde{\eta}^2 &= \widetilde{\delta}_4\eta^2
\end{aligned}
$$

We have stated that if either $C1 \neq 0$ or $C2 \neq 0$ or $C3 \neq 0$ are satisfied,then

$$\alpha = \widetilde{\alpha}, \qquad \beta = \widetilde{\beta}, \qquad \gamma = \widetilde{\gamma}, \qquad \eta = \widetilde{\eta}, \qquad \upsilon = \widetilde{\upsilon},$$

which implies $\delta_i = \widetilde{\delta}_i$ for $i = 1, \ldots, 12$.

□

4.D The cases where $C_1 \cdot C_2 \cdot C_3 = 0$

In this Appendix we describe a procedure to obtain equivalent $BMAP_2(2)$ to a given one, in the case that the general condition $C_1 \cdot C_2 \cdot C_3 \neq 0$ is not satisfied and under the assumption that $\tau \neq 0$, being τ defined in (4.10). These cases can be summarized as

1. $C_2 \cdot C_3 \neq 0$ and $C_1 = 0$,

2. $C_1 \cdot C_3 \neq 0$ and $C_2 = 0$,

3. $C_1 \cdot C_2 \neq 0$ and $C_3 = 0$,

4. $C_2 = C_3 = 0$ and $C_1 \neq 0$,

5. $C_1 = C_3 = 0$ and $C_2 \neq 0$,

6. $C_1 = C_2 = 0$ and $C_3 \neq 0$.

The case where $C_1 = C_2 = C_3 = 0$ is not considered since $\tau = C_2 + C_3$ is assumed to be different from zero. We claim that cases 4-5-6 are not possible in practice. It can be easily seen that if $C_i = 0$, then either $C_j \cdot C_k \neq 0$ or $C_j = C_k = 0$, for $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$. Therefore, there does not exist any $BMAP_2(2)$ such that $C_j = C_k = 0$ and $C_i \neq 0$, for $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$.

We focus now on the cases 1-3, under the assumption that $\tau \neq 0$. The solution to the system of equation (4.17) for each case is given by the next three results.

Proposition 4.3. *Let $\mathcal{B}$ be a $BMAP_2(2)$ as in (4.6) with underlying MAP_2, $\mathcal{M}$, as in (4.7). Assume that*

A1. $C_2 \cdot C_3 \neq 0$, $C_1 = 0$,

A2. $\mathcal{M}$ *satisfies* $\tau \neq 0$.

Let $\tilde{u} < 0$ and $\tilde{r} > 0$, and let $\tilde{x}(\tilde{u}, \tilde{r})$, $\tilde{y}(\tilde{u}, \tilde{r})$, $\tilde{v}(\tilde{u}, \tilde{r})$, $\tilde{w}(\tilde{u}, \tilde{r})$ be defined as in (4.18), and

$$\begin{aligned}
\tilde{q}(\tilde{u}, \tilde{r}) &= f_1(\widetilde{\phi}, \tilde{x}, \tilde{y}, D_0, D_1, D_2), \\
\tilde{m}(\tilde{u}, \tilde{r}) &= f_2(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2), \\
\tilde{n}(\tilde{u}, \tilde{r}) &= f_3(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2), \\
\tilde{t}(\tilde{u}, \tilde{r}) &= f_4(\widetilde{\phi}, \tilde{x}, \tilde{y}, \tilde{q}, D_0, D_1, D_2),
\end{aligned} \tag{4.34}$$

for specific values of functions f_1, f_2, f_3 and f_4. Then, the set of values $\{\tilde{u}, \tilde{r}, \tilde{x}(\tilde{u}, \tilde{r}), \tilde{y}(\tilde{u}, \tilde{r}), \tilde{v}(\tilde{u}, \tilde{r}), \tilde{w}(\tilde{u}, \tilde{r}), \tilde{m}(\tilde{u}, \tilde{r}), \tilde{q}(\tilde{u}, \tilde{r}), \tilde{n}(\tilde{u}, \tilde{r}), \tilde{t}(\tilde{u}, \tilde{r})\}$ solves the system of equations given by (4.17).

Proposition 4.4. *Let $\mathcal{B}$ be a $BMAP_2(2)$ as in (4.6) with underlying MAP_2, $\mathcal{M}$, as in (4.7). Assume that*

A1. $C_1 \cdot C_3 \neq 0$, $C_2 = 0$,

A2. $\mathcal{M}$ *satisfies* $\tau \neq 0$.

Let $\tilde{u} < 0$ *and* $\tilde{r} > 0$, *and let* $\tilde{x}(\tilde{u},\tilde{r})$, $\tilde{y}(\tilde{u},\tilde{r})$, $\tilde{v}(\tilde{u},\tilde{r})$, $\tilde{w}(\tilde{u},\tilde{r})$ *be defined as in (4.18), and*

$$\begin{aligned}
\tilde{q}(\tilde{u},\tilde{r}) &= g_1(\widetilde{\phi},\tilde{x},\tilde{y},D_0,D_1,D_2),\\
\tilde{m}(\tilde{u},\tilde{r}) &= g_2(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),\\
\tilde{n}(\tilde{u},\tilde{r}) &= g_3(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),\\
\tilde{t}(\tilde{u},\tilde{r}) &= g_4(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),
\end{aligned} \tag{4.35}$$

for specific values of functions g_1, g_2, g_3 *and* g_4. *Then, the set of values* $\{\tilde{u},\tilde{r},\tilde{x}(\tilde{u},\tilde{r}),\tilde{y}(\tilde{u},\tilde{r}),\tilde{v}(\tilde{u},\tilde{r}),\tilde{w}(\tilde{u},\tilde{r}),\tilde{m}(\tilde{u},\tilde{r}),\tilde{q}(\tilde{u},\tilde{r}),\tilde{n}(\tilde{u},\tilde{r}),\tilde{t}(\tilde{u},\tilde{r})\}$ *solves the system of equations given by (4.17).*

Proposition 4.5. *Let* $\mathcal{B}$ *be a BMAP*$_2(2)$ *as in (4.6) with underlying MAP*$_2$, $\mathcal{M}$, *as in (4.7). Assume that*

A1. $C_1 \cdot C_2 \neq 0$, $C_3 = 0$,

A2. $\mathcal{M}$ *satisfies* $\tau \neq 0$.

Let $\tilde{u} < 0$ *and* $\tilde{r} > 0$, *and let* $\tilde{x}(\tilde{u},\tilde{r})$, $\tilde{y}(\tilde{u},\tilde{r})$, $\tilde{v}(\tilde{u},\tilde{r})$, $\tilde{w}(\tilde{u},\tilde{r})$ *be defined as in (4.18), and*

$$\begin{aligned}
\tilde{q}(\tilde{u},\tilde{r}) &= q_1(\widetilde{\phi},\tilde{x},\tilde{y},D_0,D_1,D_2),\\
\tilde{m}(\tilde{u},\tilde{r}) &= q_2(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),\\
\tilde{n}(\tilde{u},\tilde{r}) &= q_3(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),\\
\tilde{t}(\tilde{u},\tilde{r}) &= q_4(\widetilde{\phi},\tilde{x},\tilde{y},\tilde{q},D_0,D_1,D_2),
\end{aligned} \tag{4.36}$$

for specific values of functions q_1, q_2, q_3 *and* q_4. *Then, the set of values* $\{\tilde{u},\tilde{r},\tilde{x}(\tilde{u},\tilde{r}),\tilde{y}(\tilde{u},\tilde{r}),\tilde{v}(\tilde{u},\tilde{r}),\tilde{w}(\tilde{u},\tilde{r}),\tilde{m}(\tilde{u},\tilde{r}),\tilde{q}(\tilde{u},\tilde{r}),\tilde{n}(\tilde{u},\tilde{r}),\tilde{t}(\tilde{u},\tilde{r})\}$ *solves the system of equations given by (4.17).*

Remark 4.5. *The analogous to Remark 4.2 applies to Propositions 3-5.*

Remark 4.6. *Closed-form expressions for* f_i, g_i *and* q_i, *for* $i = 1,2,3,4$ *can be found at*

https://sites.google.com/site/joavrc/software

Remark 4.7. *Similarly as happened for the general case $C_1 \cdot C_2 \cdot C_3 \neq 0$, the set of values in (4.34), (4.35) and (4.36) solves the equality of Laplace transforms (4.5) for $n = 1, 2$, and according to the proof of Corollary 4.1 this is generalized to all $n \geq 1$. In addition, in order to obtain a real equivalent* BMAP$_2$(2)*, Proposition 4.2 also applies.*

Remark 4.8. *The analogous to Remark 4.4 applies to Propositions 3-5.*

4.E Theorem 4.1 of Ramírez-Cobo et al. [87]

Consider the two-state *MAP* or *MAP*$_2$, characterized by $\mathcal{M} \equiv \{\boldsymbol{\theta}, D_0, D_1\}$ where

$$\boldsymbol{\theta} = (\theta,\ 1-\theta), \quad D_0 = \begin{pmatrix} x & y \\ z & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} w & -x-y-w \\ v & -z-u-v \end{pmatrix}, \tag{4.37}$$

and

$$x = -\lambda_1, \quad y = \lambda_1 p_{120}, \quad w = \lambda_1 p_{111},$$
$$z = \lambda_2 p_{210}, \quad u = -\lambda_2, \quad v = \lambda_2 p_{211}.$$

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

$$\phi = \frac{wz - vx}{wz - vx - zy - vy + xu + wu}.$$

Theorem 4.3 (Theorem 4.1 of Ramírez-Cobo et al. [87])**.** *Consider a MAP$_2$, $\mathcal{M}$ as in (4.37), and define*

$$\begin{aligned}
\varepsilon_1 &= -x, \\
\varepsilon_2 &= \frac{u-x}{2}, \\
\varepsilon_3 &= \frac{z(1-\phi)}{\phi}, \\
\varepsilon_4 &= \frac{(u-x)+\sqrt{(x-u)^2+4zy}}{2}, \\
\varepsilon_5 &= -\frac{z}{v}(z+u+v), \\
\varepsilon_6 &= -\frac{z}{2v}\left[(u+v+z+w)-\sqrt{(u+v+z+w)^2+4v(-w-y-x)}\right].
\end{aligned}$$

Let ε be chosen from

$$0 < \varepsilon < \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6\}, \quad \text{if } x < u, \tag{4.38}$$

$$0 < \varepsilon < \min\{\varepsilon_1, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6\}, \quad \text{if } x = u, \tag{4.39}$$

and set $\tilde{u} \equiv u - \varepsilon$ and $\tilde{z} \equiv z + \varepsilon$. Then there exist an infinite number of MAP$_2$s, $\widetilde{\mathcal{M}}$, given by $\mathcal{F} = \{\tilde{u}, \tilde{z}, \tilde{x}(\tilde{u}, \tilde{z}), \tilde{y}(\tilde{u}, \tilde{z}), \tilde{v}(\tilde{u}, \tilde{z}), \tilde{w}(\tilde{u}, \tilde{z})\}$, where $\tilde{x}(\tilde{u}, \tilde{z})$, $\tilde{y}(\tilde{u}, \tilde{z})$, $\tilde{v}(\tilde{u}, \tilde{z})$, and $\tilde{w}(\tilde{u}, \tilde{z})$ are defined in Proposition 4.2 of Ramírez-Cobo et al. [87], such that the equality of the Laplace transforms holds for $n = 1, 2$.

4.F Proof of Proposition 4.2

We prove here that the set $\mathcal{F}$ in Proposition 4.2 provides feasible solutions to the problem of equivalent $BMAP_2(2)$s. Assume first that $x < u$. Let κ be defined as in (4.23), that is,

$$0 < \kappa < min\{\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6, \kappa_7, \kappa_8\}.$$

First, we prove that $\min\{\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6, \kappa_7, \kappa_8\} > 0$. It is straightforward to check that

$$min\{\kappa_1, \kappa_2, \kappa_3, \kappa_4\} > 0.$$

Also, $\kappa_5 = -\frac{r}{t}(r + u + v + q + t) > 0$, since $r, t > 0$ and $-r - u - v - q - t > 0$. Next, $\kappa_6 > 0$ since $u - x > 0$, and $\kappa_7 > 0$ since $(q - w) < \sqrt{(q - w)^2 + 4vm}$, because $v, m > 0$. Finally, $\kappa_8 > 0$ since $-\dfrac{r}{2t} < 0$, $-y - x - w - m - n > 0$ and therefore,

$$(u + v + q + n + t + r)^2 < (u + v + q + n + t + r)^2 + 4t(-y - x - w - m - n).$$

Then,

$$\tilde{u} = u - \kappa < 0 \quad \text{and} \quad \tilde{r} = r + \kappa > 0.$$

Moreover, since $\kappa < \kappa_2 = \dfrac{u - x}{2}$, this assures that $\tilde{x} < \tilde{u}$, and thus the parameterization of $\widetilde{\mathcal{M}}$ is different from that of $\mathcal{M}$ with permuted states.

Since $\kappa < \kappa_3 = \dfrac{r(1 - \phi)}{\phi}$, then

$$0 \leq \tilde{\phi} \equiv \frac{(r + \kappa)\phi}{r} \leq 1.$$

Also, from $\kappa < \kappa_4 = \dfrac{rq}{v}$,

$$\tilde{q} \equiv \frac{qr - v\kappa}{r} > 0,$$

is obtained. Next,

$$\frac{(u - x) - \sqrt{(x - u)^2 + 4ry}}{2} < 0 < \kappa < \frac{(u - x) + \sqrt{(x - u)^2 + 4ry}}{2} = \kappa_6,$$

implies that

$$\tilde{y}(\tilde{u},\tilde{r}) \equiv \frac{-(\kappa^2 + (x-u)\kappa - ry)}{r+\kappa} > 0,$$

and

$$\frac{r}{2v}\left[(q-w) - \sqrt{(q-w)^2 + 4vm}\right] < 0 < \kappa < \frac{r}{2v}\left[(q-w) + \sqrt{(q-w)^2 + 4vm}\right] = \kappa_7,$$

implies that

$$\tilde{m}(\tilde{u},\tilde{r}) \equiv \frac{-(v\kappa^2 - r(q-w)\kappa - mr^2)}{r(r+\kappa)} > 0.$$

In addition,

$$\tilde{w}(\tilde{u},\tilde{r}) \equiv \frac{wr + v\kappa}{r} > 0, \qquad \tilde{v}(\tilde{u},\tilde{r}) \equiv \frac{v(r+\kappa)}{r} > 0,$$
$$\tilde{t}(\tilde{u},\tilde{r}) \equiv \frac{t(r+\kappa)}{r} > 0, \qquad \tilde{n}(\tilde{u},\tilde{r}) \equiv \frac{nr + t\kappa}{r} > 0.$$

It remains to prove that $-\tilde{r} - \tilde{u} - \tilde{v} - \tilde{q} - \tilde{t} > 0$ and $-\tilde{x} - \tilde{y} - \tilde{w} - \tilde{m} - \tilde{n} > 0$. It is easy to check that

$$\begin{aligned} -\tilde{r} - \tilde{u} - \tilde{v} - \tilde{q} - \tilde{t} &= -r - u - \frac{v(r+\kappa)}{r} - \frac{qr - v\kappa}{r} - \frac{t(r+\kappa)}{r} \\ &= -r - u - \frac{r(v+q) + t(r+\kappa)}{r}, \end{aligned}$$

which is positive if and only if $\kappa < \kappa_5 = -\frac{r}{t}(r+u+v+q+t)$. Finally, an easy computation shows that $-\tilde{x} - \tilde{y} - \tilde{w} - \tilde{m} - \tilde{n} > 0$ is equivalent to

$$-\kappa - x > \frac{-(\kappa^2 + (x-u)\kappa - ry)}{r+\kappa} + \frac{-(v\kappa^2 - r(q-w)\kappa - mr^2)}{r(r+\kappa)} + \frac{(nr + t\kappa) + (wr + v\kappa)}{r},$$

which holds if and only if $\kappa \in (r_1, r_2)$ where

$$\begin{aligned} r_1 &= -\frac{r}{2t}\Big[(u+v+q+n+t+r) \\ &\quad + \sqrt{(u+v+q+n+t+r)^2 + 4t(-y-w-m-n-x)}\Big] < 0, \\ r_2 &= -\frac{r}{2t}\Big[(u+v+q+n+t+r) \\ &\quad - \sqrt{(u+v+q+n+t+r)^2 + 4t(-y-w-m-n-x)}\Big] = \kappa_8 > 0. \end{aligned}$$

Now, assume that $x = u$. Then, let κ be defined as in (4.24), where in this case, $\kappa_6 \equiv \sqrt{ry}$. Then,

$$
\begin{aligned}
\tilde{u} &= u - \kappa < 0, \\
\tilde{r} &= r + \kappa > 0, \\
\tilde{x} &= x + \kappa < 0 \quad (\text{since } \kappa < -x), \\
\tilde{y} &= \frac{ry - \kappa^2}{r + \kappa} > 0 \quad (\text{since } \kappa < \sqrt{ry}), \\
\tilde{w} &= \frac{wr + v\kappa}{r} > 0, \\
\tilde{v} &= \frac{v(r + \kappa)}{r} > 0, \\
\tilde{n} &= \frac{nr + t\kappa}{r} > 0, \\
\tilde{t} &= \frac{t(r + \kappa)}{r} > 0,
\end{aligned}
$$

and $\tilde{q}$, $\tilde{m}$, $\tilde{\phi} \in [0,1]$, $-\tilde{r} - \tilde{u} - \tilde{v} - \tilde{q} - \tilde{t} > 0$ and $-\tilde{x} - \tilde{y} - \tilde{w} - \tilde{m} - \tilde{n} > 0$ follow from the assumptions $\kappa < \kappa_4$, $\kappa < \kappa_7$, $\kappa < \kappa_3$, $\kappa < \kappa_5$ and $\kappa < \kappa_8$, respectively.

□

Chapter 5

Dependence patterns of the *BMAP*

The Batch Markovian Arrival Process allows for correlated batch event occurrences and dependent inter-event times. Therefore it is considered as a model in contexts where dependent data is observed. Most works regarding the theoretical aspect of the correlation structure are focused on special cases of the MAP, specifically, the two-state MAP, see Heindl et al. [41], Casale et al. [13], Ramírez-Cobo and Carrizosa [92] and Hervé and Ledoux [42], where different analysis of the inter-event times correlation function are carried out, such as correlation bounds, behaviors and characterizations. The main result found is that auto-correlation function of the inter-event times for the stationary MAP_m decreases geometrically.

The auto-correlation function for a sequence of inter-event times of a $BMAP$ has a known closed-form, as it can be seen in Chakravarthy [15]. However, the structure of the correlation of the batch arrivals has not been studied in detail in the literature. In this chapter we study some theoretical properties of the correlation functions of the inter-event times and batches sizes for the stationary $BMAP_2(2)$. We prove that both auto-correlation functions decrease geometrically to zero. In addition, we provide a characterization of the correlation functions for the general $BMAP_m(k)$ case.

This chapter is organized as follows. Section 5.1 contains some preliminary results concerning the correlation functions of the inter-event times and batches sizes. Section 5.2 generalizes the theoretical characterization of the auto-correlation functions of the general $BMAP_m(k)$, for $k \geq 2$. In Section 5.3 we provide closed-form functions for the auto-correlation structures for the $BMAP_2(2)$, and provide important theoretical results and patterns. Finally, in Section 5.4 some conclusions of this chapter are given.

5.1 Preliminaries

In this section we recall the auto-correlation formulas for the inter-event times and batch sizes, introduced in the previous chapter, in (4.3) and (4.4) respectively, which are the starting point of the contributions in this chapter.

The auto-correlation function of the time between the $(n-1)$-th and n-th event occurrences, T_n, in the stationary version, is given by

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

where $l \geq 1$ represents the time lag, $D = D_1 + D_2$, and

$$\begin{aligned} \mu_T &= \boldsymbol{\phi}(-D_0)^{-1}\mathbf{e}, \\ \sigma_T^2 &= \mu_T(2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e} - \mu_T). \end{aligned}$$

Recall that $\boldsymbol{\pi} = \boldsymbol{\pi}_{BMAP}$ is the stationary probability vector of the underlying Markov process $J(t)$, and $\boldsymbol{\phi} = \boldsymbol{\phi}_{BMAP}$ is defined in (1.14) as

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

Equivalently,

$$\rho_T(l) = \frac{\left(\boldsymbol{\pi} \left[(-D_0)^{-1}D\right]^l (-D_0)^{-1}\mathbf{e} - \mu_T\right)}{2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e} - \mu_T}. \tag{5.1}$$

Let B_n denote the size of the n-th batch event occurrence in the stationary version of the $BMAP_m(k)$. Then, the auto-correlation function is given by

$$\rho_B(l) = \rho(B_1, B_{l+1}) = \frac{\boldsymbol{\phi}(-D_0)^{-1}D_1^\star \left[(-D_0)^{-1}D\right]^{l-1} (-D_0)^{-1}D_1^\star\mathbf{e} - (\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e})^2}{\boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e} - (\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e})^2}, \tag{5.2}$$

where $l \geq 1$ represents the time lag, and $D_1^\star = D_1 + 2D_2$ and $D_2^\star = D_1 + 4D_2$, and

$$\begin{aligned} \mu_B &= \boldsymbol{\phi}(-D_0)^{-1}(D_1 + 2D_2)\mathbf{e}, \\ \sigma_B^2 &= \boldsymbol{\phi}(-D_0)^{-1}(D_1 + 4D_2)\mathbf{e} - \left(\boldsymbol{\phi}(-D_0)^{-1}(D_1 + 2D_2)\mathbf{e}\right)^2. \end{aligned}$$

In the following, we will try to obtain information from (5.1) and (5.2).

5.2 Dependence structure of the $\boldsymbol{BMAP}_m(k)$, $m \geq 3$

This section generalizes previous results of the correlation structure of the inter-event times for the stationary MAP_2 and introduce new results on the correlation structure of the batch size.

Theorem 5.1. *Consider a $BMAP_m(k)$, for $k \geq 2$, and let $\rho_T(l)$ and $\rho_B(l)$ denote the auto-correlation function of the inter-event times distribution and batch sizes respectively. Then*

$$\begin{aligned}\rho_T(l) &= \sum_{i=2}^{m} p_{m,k,i}(T) q_{k,i}^{l},\\ \rho_B(l) &= \sum_{i=2}^{m} p_{m,k,i}(B) q_{k,i}^{l-1},\end{aligned}$$

where $q_{k,i}$, for $i = 2, \dots, m$ are the eigenvalues of the matrix $(-D_0)^{-1}D$ different from 1, such that $|q_{k,i}| < 1$, and $p_{m,k,i}(T)$ and $p_{m,k,i}(B)$ are derived after calculations.

In addition,

$$\begin{aligned}|\rho_T(l)| &\geq |\rho_T(l+1)|, \quad \text{for all } l \geq 1 \text{ and } \lim_{l\to\infty} \rho_T(l) = 0,\\ |\rho_B(l)| &\geq |\rho_B(l+1)|, \quad \text{for all } l \geq 1 \text{ and } \lim_{l\to\infty} \rho_B(l) = 0.\end{aligned}$$

Proof. Note that from (5.1) and (5.2), the matrix $[(-D_0)^{-1}D]$ is the stochastic matrix $P^\star$, therefore, for the general m-state case, $[(-D_0)^{-1}D]^l$ has spectral decomposition,

$$\left[(-D_0)^{-1}D\right]^l = Q\Delta^l Q^{-1},$$

where $\Delta = diag(q_{k,1}, q_{k,2}, \dots, q_{k,m})$ are the eigenvalues, Q is the right eigenvector and Q^{-1} is the left eigenvector of $(-D_0)^{-1}D$.

For proving the result, we need the classical Perron-Frobenious theorem (for its proof and many applications, we refer the reader to MacCluer [64] or Lawler [55], pages 16-17).

Theorem 5.2. *(Perron-Frobenious theorem) Let A be a $m \times m$ stochastic matrix. Then the following hold:*

i) 1 is a simple eigenvalue of A.

ii) The absolute value of any other eigenvalue is strictly less than 1, that is, all eigenvalues $q_{k,i} \neq 1$ satisfy, $|q_{k,i}| < 1$.

Since,

$$\boldsymbol{\pi} P^{\star} = \boldsymbol{\pi} \Leftrightarrow \boldsymbol{\pi} = \boldsymbol{\phi}, \qquad \text{and } P^{\star}\mathbf{e} = \mathbf{e},$$

we have that the left eigenvector associated to the eigenvalue 1 is $\boldsymbol{\phi}$, and the right eigenvector associated to 1 is $\mathbf{e}$.

Then,

$$\left[(-D_0)^{-1}D\right]^l = [\mathbf{e} \quad \nu^{(2)} \quad \nu^{(3)} \dots \nu^{(m)}]\Delta^l[\boldsymbol{\phi} \quad \omega^{(2)} \quad \omega^{(3)} \dots \omega^{(m)}]^{\intercal},$$

where $\nu^{(i)}$ and $\omega^{(i)}$ are the columns and rows of the right and left eigenvectors respectively, so $\omega^{(i)}\nu^{(i)} = 1$, and $\Delta^l = diag(1, q^l_{k,2}, q^l_{k,3} \dots, q^l_{k,m})$, so

$$\begin{aligned}\left[(-D_0)^{-1}D\right]^l &= \mathbf{e}\boldsymbol{\phi} + \nu^{(2)}\omega^{(2)}q^l_{k,2} + \nu^{(3)}\omega^{(3)}q^l_{k,3} + \dots \\ &= \mathbf{e}\boldsymbol{\phi} + \sum_{i=2}^{m} \nu^{(i)}\omega^{(i)}q^l_{k,i}. \end{aligned} \tag{5.3}$$

Now, for simplicity purposes, denote the numerators of (5.1) and (5.2) as

$$\begin{aligned}\tau^m_T(l) &= \boldsymbol{\pi}\left[(-D_0)^{-1}D\right]^l(-D_0)^{-1}\mathbf{e} - \mu_T, \\ \tau^m_B(l) &= \boldsymbol{\phi}(-D_0)^{-1}D^{\star}_1\left[(-D_0)^{-1}D\right]^{l-1}(-D_0)^{-1}D^{\star}_1\mathbf{e} - (\boldsymbol{\phi}(-D_0)^{-1}D^{\star}_1\mathbf{e})^2.\end{aligned}$$

Recall that $\mu_T = \boldsymbol{\phi}(-D_0)^{-1}\mathbf{e}$, then, substituting (5.3) in $\tau^m_T(l)$, we obtain

$$\begin{aligned}\tau^m_T(l) &= \boldsymbol{\pi}\left[(-D_0)^{-1}D\right]^l(-D_0)^{-1}\mathbf{e} - \mu_T \\ &= \boldsymbol{\pi}\left[\mathbf{e}\boldsymbol{\phi} + \sum_{i=2}^{m}\nu^{(i)}\omega^{(i)}q^l_{k,i}\right](-D_0)^{-1}\mathbf{e} - \mu_T \\ &= \boldsymbol{\pi}\mathbf{e}\boldsymbol{\phi}(-D_0)^{-1}\mathbf{e} + \boldsymbol{\pi}\left[\sum_{i=2}^{m}\nu^{(i)}\omega^{(i)}q^l_{k,i}\right](-D_0)^{-1}\mathbf{e} - \mu_T \\ &= \sum_{i=2}^{m}\left(\boldsymbol{\pi}A_i(-D_0)^{-1}\mathbf{e}\right)q^l_{k,i},\end{aligned}$$

where

$$A_i = \nu^{(i)}\omega^{(i)}.$$

Analogously,

$$\tau^m_B(l) = \sum_{i=2}^{m}\left(\boldsymbol{\phi}(-D_0)^{-1}D^{\star}_1A_i(-D_0)^{-1}D^{\star}_1\mathbf{e}\right)q^{l-1}_{k,i}.$$

Then,

$$\begin{aligned}\rho_T(l) &= \frac{\tau_T^m(l)}{2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e}-\mu_T} \\ &= \frac{\sum_{i=2}^m \left(\boldsymbol{\pi}A_i(-D_0)^{-1}\mathbf{e}\right)q_{k,i}^l}{2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e}-\mu_T} \\ &= \sum_{i=2}^m \frac{\left(\boldsymbol{\pi}A_i(-D_0)^{-1}\mathbf{e}\right)q_{k,i}^l}{2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e}-\mu_T}.\end{aligned}$$

And

$$\begin{aligned}\rho_B(l) &= \frac{\tau_B^m(l)}{\boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e}-\left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e}\right)^2} \\ &= \frac{\sum_{i=2}^m \left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star A_i(-D_0)^{-1}D_1^\star\mathbf{e}\right)q_{k,i}^{l-1}}{\boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e}-\left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e}\right)^2} \\ &= \sum_{i=2}^m \frac{\left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star A_i(-D_0)^{-1}D_1^\star\mathbf{e}\right)q_{k,i}^{l-1}}{\boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e}-\left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e}\right)^2}.\end{aligned}$$

Therefore, the auto-correlation of the inter-event times for the general $BMAP_m(k)$ for $k \geq 2$ can be written as

$$\rho_T(l) = \sum_{i=2}^m p_{m,k,i}(T)q_{k,i}^l \tag{5.4}$$

where

$$p_{m,k,i}(T) = \left(\frac{\boldsymbol{\pi}A_i(-D_0)^{-1}\mathbf{e}}{2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e}-\mu_T}\right)$$

And for the batches,

$$\rho_B(l) = \sum_{i=2}^m p_{m,k,i}(B)q_{k,i}^{l-1} \tag{5.5}$$

where

$$p_{m,k,i}(B) = \left(\frac{\boldsymbol{\phi}(-D_0)^{-1}D_1^\star A_i(-D_0)^{-1}D_1^\star\mathbf{e}}{\boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e}-\left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e}\right)^2}\right)$$

Since $|q_{k,i}| < 1$ we conclude that $\rho_T(l)$ and $\rho_B(l)$ are also decreasing sequences. □

The complexity of the auto-correlation functions increases as the number of states, m, of the process increases. Therefore, we expect to obtain richer correlation structures for higher order *BMAP*s.

5.3 Dependence structures of the $\boldsymbol{BMAP}_2(k)$, $k \geq 2$

In the this section we tackled the auto-correlation structure for the two-state case. This section aims to provide a correlation structure for the general $BMAP_m(k)$, for $k \geq 2$.

Recall that the stationary $BMAP_2(2)$ is represented by $\mathcal{B} = \{D_0, D_1, D_2\}$ where (see equation (4.6) from Chapter 4)

$$D_0 = \begin{pmatrix} x & y \\ r & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} w & m \\ v & q \end{pmatrix}, \quad D_2 = \begin{pmatrix} n & -x-y-w-m-n \\ t & -r-u-v-q-t \end{pmatrix},$$

where without loss of generality it is assumed that $x \leq u$, and

$$x = -\lambda_1, \quad y = \lambda_1 p_{120}, \quad w = \lambda_1 p_{111}, \quad m = \lambda_1 p_{121}, \quad n = \lambda_1 p_{112},$$

$$r = \lambda_2 p_{210}, \quad u = -\lambda_2, \quad v = \lambda_2 p_{211}, \quad q = \lambda_2 p_{221}, \quad t = \lambda_2 p_{212}.$$

We now present one of the major contributions of this chapter, where the structure of the $\rho_T(l)$ is characterized.

Lemma 5.1. *Let a $BMAP_2(2)$ determined as in (4.6), and let $\rho_T(l)$ denote the auto-correlation function of the inter-event times distribution. Then*

$$\rho_T(l) = p(T)q^l, \tag{5.6}$$

for some $p(T)$ and $|q| < 1$. In addition,

$$|\rho_T(l)| \geq |\rho_T(l+1)|, \quad \textit{for all } l \geq 1 \textit{ and } \lim_{l\to\infty} \rho_T(l) = 0.$$

Proof. Let ω_l denote the numerator in (5.1)

$$\omega_l = \boldsymbol{\pi} \left[(-D_0)^{-1}D\right]^l (-D_0)^{-1}\mathbf{e} - \mu_T. \tag{5.7}$$

And, let ω denote the denominator in (5.1)

$$\omega = 2\boldsymbol{\pi}(-D_0)^{-1}\mathbf{e} - \mu_T,$$

Then, after calculations, ω may be written in terms of the model parameters as

$$\omega = \frac{\kappa_1 + \kappa_2}{(ry-ux)(n-r-t-v+w+x)(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy)}. \tag{5.8}$$

where

$$\begin{aligned}\kappa_1 &= 2(x-r)(n+w+x)(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy)\\ &\quad +2(y-u)(r+t+v)(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy),\\ \kappa_2 &= (ry-ux)(n-r-t-v+w+x)^2.\end{aligned}$$

Now, consider the expression (5.7). It can be seen that $(-D_0)^{-1}D$ has spectral decomposition

$$(-D_0)^{-1}D = Q\Delta Q^{-1},$$

where Δ is a diagonal matrix whose components are the eigenvalues of $(-D_0)^{-1}D$ and Q is the corresponding eigenvectors matrix, given by

$$Q = \begin{pmatrix} 1 & \dfrac{ry-ux-nu-uw+ty+vy}{nr+rw-tx-vx} \\ 1 & 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{(nr+nu+rw-tx+uw-ty-vx-vy)}{ry-ux} \end{pmatrix}. \tag{5.9}$$

Therefore,

$$\left[(-D_0)^{-1}D\right]^l = Q\Delta^l Q^{-1}, \tag{5.10}$$

where

$$\Delta^l = \begin{pmatrix} 1 & 0 \\ 0 & \left[\dfrac{(nr+nu+rw-tx+uw-ty-vx-vy)}{ry-ux}\right]^l \end{pmatrix}.$$

Then, ω_l may be written in terms of the model parameters as

$$\omega_l = \frac{(\kappa_3+\kappa_4)(nr+nu+rw-tx+uw-ty-vx-vy)^l(-r-u+x+y)}{(ry-ux)^l(ry-ux)(n-r-t-v+w+x)(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy)}, \tag{5.11}$$

where

$$\begin{aligned}\kappa_3 &= n^2r+2nrw+nrx+unr-ntx+unt-nvx+unv-yr^2-2yrt-2yrv+rw^2+rwx,\\ \kappa_4 &= urw+urx-yt^2-2ytv-twx+utw-tx^2+utx-yv^2-vwx+uvw-vx^2+uvx.\end{aligned}$$

Hence, taking into account (5.8) and (5.11), it can be verified that $\rho_T(l)$ may be written in terms of the model parameters as

$$\rho_T(l) = \frac{(nr+nu+rw-tx+uw-ty-vx-vy)^l(-r-u+x+y)(\kappa_3+\kappa_4)}{(ry-ux)^l(\kappa_1+\kappa_2)}, \tag{5.12}$$

Then, $\rho_T(l) = p(T)q^l$, where

$$\begin{aligned} p(T) &= \frac{(\kappa_3+\kappa_4)(-r-u+x+y)}{(\kappa_1+\kappa_2)}, \\ q &= \frac{(nr+nu+rw-tx+uw-ty-vx-vy)}{ry-ux}. \end{aligned}$$

Note that q is an eigenvalue of $P^\star$ defined in (1.13). As $P^\star = (-D_0)^{-1}D$ is a stochastic matrix, then it has two possible eigenvalues: 1 and q, and $-1 \le q \le 1$. Then $|q| < 1$ is satisfied.

Since $|q| < 1$ we conclude that $|\rho_T(l)| \ge |\rho_T(l+1)|$ for all $l \ge 1$. □

Next, we present an analogous result for the auto-correlation function of the batch size.

Lemma 5.2. *Let a BMAP$_2(2)$ determined as in (4.6), and let $\rho_B(l)$ denote the auto-correlation function of the batch sizes. Then*

$$\rho_B(l) = p(B)q^{l-1}, \tag{5.13}$$

for some $p(B)$ and $|q| < 1$. In addition,

$$|\rho_B(l)| \ge |\rho_B(l+1)|, \quad \textit{for all } l \ge 1 \textit{ and } \lim_{l\to\infty} \rho_B(l) = 0.$$

Proof. We proceed analogous to the proof Lemma 5.1. Let τ_l denote the numerator in (5.2)

$$\tau_l = \boldsymbol{\phi}(-D_0)^{-1}D_1^\star\left[(-D_0)^{-1}D\right]^{l-1}(-D_0)^{-1}D_1^\star\mathbf{e} - (\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e})^2, \tag{5.14}$$

And, let τ denote the denominator in (5.2)

$$\tau = \boldsymbol{\phi}(-D_0)^{-1}D_2^\star\mathbf{e} - \left(\boldsymbol{\phi}(-D_0)^{-1}D_1^\star\mathbf{e}\right)^2,$$

Then, τ may be written in terms of the model parameters as

$$\tau = \frac{\epsilon_1\epsilon_2}{(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy)^2}. \tag{5.15}$$

where

$$\begin{aligned} \epsilon_1 &= -mr+nq-mt-mv+nv+qw+qx-rw-tw+vx, \\ \epsilon_2 &= mr-nq-nr+mt+mv-nu-nv-qw-qx+ry+tw+tx-uw+ty-ux+vy. \end{aligned}$$

From (5.9) and (5.10), consider the expression (5.14). It can be easily seen that

$$\left[(-D_0)^{-1}D\right]^{l-1} = Q\Delta^{l-1}Q^{-1},$$

Then, τ_k may be written in terms of the model parameters as

$$\tau_l = \frac{-\epsilon(nr+nu+rw-tx+uw-ty-vx-vy)^{l-1}(mr+mu-qx+rw-qy+uw-vx-vy)}{(ry-ux)^{l-1}(ry-ux)(nr+nu+rw-ry-tx+uw-ty+ux-vx-vy)^2}, \tag{5.16}$$

where

$$\epsilon = \epsilon_3 x^2 + (\epsilon_4+\epsilon_5)x + \epsilon_6 + \epsilon_7$$

where

$$\begin{aligned}
\epsilon_3 &= qt+qv+uv,\\
\epsilon_4 &= nqt-mv^2-v^2y-nqr-mrt-mt^2-mrv+nqv-2mtv,\\
\epsilon_5 &= 2nuv-qrw+qtw+qvw-ruw-tuw-rvy+uvw-tvy,\\
\epsilon_6 &= mnr^2-n^2qr+mr^2w+n^2uv-qrw^2-nv^2y-ruw^2-tuw^2+r^2wy+t^2wy+mnrt,\\
\epsilon_7 &= mnrv-2nqrw+mrtw+mrvw-nruw-ntuw-nrvy+nuvw-ntvy\\
&\quad +2rtwy+rvwy+tvwy.
\end{aligned}$$

Hence, taking into account (5.15) and (5.16), it is easy to check that (5.2) becomes

$$\rho_B(l) = \frac{(nr+nu+rw-tx+uw-ty-vx-vy)^{l-1}\epsilon(-mr-mu+qx-rw+qy-uw+vx+vy)}{(ry-ux)^{l-1}(ry-ux)\epsilon_1\epsilon_2},$$

Then, $\rho_B(l) = p(B)q^{l-1}$, where

$$\begin{aligned}
p(B) &= \frac{\epsilon(-mr-mu+qx-rw+qy-uw+vx+vy)}{(ry-ux)\epsilon_1\epsilon_2},\\
q &= \frac{(nr+nu+rw-tx+uw-ty-vx-vy)}{ry-ux}
\end{aligned}$$

As it was shown in Lemma 5.1, $|q|<1$ is satisfied. Therefore $\rho_B(l)$ is a decreasing sequence.

□

Remark 5.1. *Note that for any BMAP$_2$(k), for $k \geq 3$, the stochastic matrix*

$$P^\star = (-D_0)^{-1}\sum_{i=1}^{k} D_i,$$

is still a 2×2 *matrix with eigenvalues 1 and* q_k*, where* $|q_k| < 1$*. Therefore, the results for Lemmas 5.1 and 5.2 are valid for any* $BMAP_2(k)$*, for* $k \geq 3$*.*

$$\begin{aligned} \rho_T(l) &= p_k(T).q_k^l, \\ \rho_B(l) &= p_k(B).q_k^{l-1}, \end{aligned}$$

for some $p_k(T)$ *and* $p_k(B)$*, and* $|q_k| < 1$*. Therefore,* $\rho_T(l)$ *and* $\rho_B(l)$ *are decreasing sequences.*

Remark 5.2. *Expressions (5.6) and (5.13) obtained for the* $BMAP_2(2)$ *auto-correlations* ρ_T *and* ρ_B *respectively, implies that* ρ_T *and* ρ_B *can take negative or positive values for some* $l \geq 1$*, which leads to the following correlation patterns for* ρ_T*.*

- Pattern 1. If $p(T), q \geq 0$, then $\rho_T(l) \geq 0$ for all $l \geq 1$. As an example, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -0.2653 & 0.0164 \\ 0.0485 & -4.4157 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.1986 & 0.0422 \\ 0.1041 & 3.6693 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.0025 & 0.0056 \\ 0.0082 & 0.5855 \end{pmatrix},$$

  it is easy to check that $p(T) = 0.4278$ and $q = 0.7261$, and the auto-correlation function is

  $$\begin{aligned} &\rho_T(1) = 0.3106, \quad \rho_T(2) = 0.2256, \quad \rho_T(3) = 0.1638, \\ &\rho_T(4) = 0.1189, \quad \rho_T(5) = 0.0864, \quad \rho_T(6) = 0.0627. \end{aligned}$$

- Pattern 2. If $p(T) \leq 0$ and $q \geq 0$, then $\rho_T(l) \leq 0$ for all $l \geq 1$. For example, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -1.7318 & 0.7232 \\ 0.0148 & -1.7898 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0847 & 0.0159 \\ 0.1445 & 0.8245 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.9049 & 0.0031 \\ 0.7169 & 0.0890 \end{pmatrix},$$

  it is easy to check that $p(T) = -0.0534$ and $q = 0.2874$, and the auto-correlation function is

  $$\begin{aligned} &\rho_T(1) = -0.0153, \quad \rho_T(2) = -0.0044, \quad \rho_T(3) = -0.0013, \\ &\rho_T(4) = -0.0004, \quad \rho_T(5) = -0.0001, \quad \rho_T(6) = -3.0072 \times 10^{-5}. \end{aligned}$$

- Pattern 3. If $p(T) \geq 0$ and $q \leq 0$, then $\rho_T(2l) \geq 0$ and $\rho_T(2l+1) \leq 0$ for all $l \geq 1$. As an illustration, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -0.3414 & 0.0203 \\ 0.3617 & -4.6970 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0051 & 0.3080 \\ 2.7217 & 0.0159 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.0017 & 0.0063 \\ 0.7457 & 0.8519 \end{pmatrix},$$

it is easy to check that $p(T) = 0.3035$ and $q = -0.6790$, and the auto-correlation function is

$$\rho_T(1) = -0.2060, \quad \rho_T(2) = 0.1399, \quad \rho_T(3) = -0.0950,$$
$$\rho_T(4) = 0.0645, \quad \rho_T(5) = -0.0438, \quad \rho_T(6) = 0.0297.$$

- Pattern 4. If $p(T) \leq 0$ and $q < 0$, then $\rho_T(2l) \leq 0$ and $\rho_T(2l+1) \geq 0$ for all $l \geq 1$. As an example, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -1.3722 & 0.9891 \\ 0.0742 & -1.9044 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0466 & 0.2619 \\ 0.9831 & 0.0567 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.0203 & 0.0543 \\ 0.7871 & 0.0035 \end{pmatrix},$$

  it is easy to check that $p(T) = -0.1450$ and $q = -0.2188$, and the auto-correlation function is

  $$\rho_T(1) = 0.0317, \quad \rho_T(2) = -0.0069, \quad \rho_T(3) = 0.0015,$$
  $$\rho_T(4) = -0.0003 \quad \rho_T(5) = 0.0001, \quad \rho_T(6) = -1.5886 \times 10^{-5}.$$

Analogous patterns are found for $\rho_B(l)$ for $l \geq 1$. That is,

- Pattern 1. If $p(B), q \geq 0$, then $\rho_B(l) \geq 0$ for all $l \geq 1$. As an example, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -1.1791 & 0.0281 \\ 0.0462 & -1.0961 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0302 & 0.1324 \\ 0.0383 & 0.7907 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.9118 & 0.0766 \\ 0.1595 & 0.0614 \end{pmatrix},$$

  it is easy to check that $p(B) = 0.4958$ and $q = 0.5897$, and the auto-correlation function is

  $$\rho_B(1) = 0.4958, \quad \rho_B(2) = 0.2924, \quad \rho_B(3) = 0.1724,$$
  $$\rho_B(4) = 0.1017, \quad \rho_B(5) = 0.0600, \quad \rho_B(6) = 0.0354.$$

- Pattern 2. If $p(B) \leq 0$ and $q \geq 0$, then $\rho_B(l) \leq 0$ for all $l \geq 1$. For example, consider the $BMAP_2(2)$,

  $$D_0 = \begin{pmatrix} -1.8815 & 0.0719 \\ 0.0551 & -1.3879 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.9732 & 0.8188 \\ 0.0216 & 0.6550 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.0116 & 0.0060 \\ 0.6001 & 0.0562 \end{pmatrix},$$

  it is easy to check that $p(B) = -0.2396$ and $q = 0.0720$, and the auto-correlation function is

  $$\rho_B(1) = -0.2396, \quad \rho_B(2) = -0.0172, \quad \rho_B(3) = -0.0012,$$
  $$\rho_B(4) = -0.0001, \quad \rho_B(5) = -6.4348 \times 10^{-6}, \quad \rho_B(6) = -4.6324 \times 10^{-7}.$$

- Pattern 3. If $p(B) \geq 0$ and $q \leq 0$, then $\rho_B(2l) \leq 0$ and $\rho_B(2l+1) \geq 0$ for all $l \geq 1$. As an illustration, consider the $BMAP_2(2)$,

$$D_0 = \begin{pmatrix} -1.8499 & 0.0233 \\ 0.3029 & -1.7062 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.0172 & 0.7854 \\ 0.9346 & 0.4201 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.9496 & 0.0743 \\ 0.0485 & 0.0001 \end{pmatrix},$$

  it is easy to check that $p(B) = 0.2352$ and $q = -0.1394$, and the auto-correlation function is

$$\rho_B(1) = 0.2352, \quad \rho_B(2) = -0.0328, \quad \rho_B(3) = 0.0046,$$
$$\rho_B(4) = -0.0006, \quad \rho_B(5) = 0.0001, \quad \rho_B(6) = -1.2373 \times 10^{-5}.$$

- Pattern 4. If $p(B) \leq 0$ and $q < 0$, then $\rho_B(2l) \geq 0$ and $\rho_B(2l+1) \leq 0$ for all $l \geq 1$. As an example, consider the $BMAP_2(2)$,

$$D_0 = \begin{pmatrix} -1.5013 & 0.1198 \\ 0.0836 & -1.1887 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.3441 & 0.9929 \\ 0.0759 & 0.0240 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.0420 & 0.0026 \\ 0.9269 & 0.0783 \end{pmatrix},$$

  it is easy to check that $p(B) = -0.4754$ and $q = -0.5402$, and the auto-correlation function is

$$\rho_B(1) = -0.4754, \quad \rho_B(2) = 0.2568, \quad \rho_B(3) = -0.1387,$$
$$\rho_B(4) = 0.0749, \quad \rho_B(5) = -0.0405, \quad \rho_B(6) = 0.0219.$$

5.4 Chapter summary

This chapter extends important properties of the auto-correlation function of the inter-event times of the stationary $BMAP_2(2)$, and introduces new results for the auto-correlation function of the batch sizes. We provide a characterization of both auto-correlation functions in terms of the eigenvalues of $P^\star$, and prove that they both decrease geometrically as the time lag increases. In addition, four behavior patterns are distinguished for both correlation functions, illustrated with numerical examples. Also, the characterization of the auto-correlation functions has been extended for the general $BMAP_m(k)$ case.

Chapter 6

Conclusions and future work

In this dissertation we have studied important properties of the point process that captures the statistical features of dependent and not identical distributed inter-event times (MAP), as well as possible correlates batch arrivals ($BMAP$). These processes have been proved to be manageable and tractable models. And therefore are a good alternative to model non-exponential type inter-event times. The main contributions of this dissertation are the following,

- We have studied the non-stationary Markovian arrival process, in particular its two-state case, MAP_2, and we characterize its canonical representation. Specifically,
  - ✓ We have defined when a non-stationary MAP_2 is identifiable in terms of its Laplace-Stieltjes transform for all $n \geq 1$ given that the inter-event times are observed.
  - ✓ We show that the non-stationary MAP_2 is a nonidentifiable process, which is also a valid result for non-stationary MAP's of order m.
  - ✓ We consider a moments matching approach in order to provide a procedure that shows how to build an equivalent non-stationary MAP_2 for any given fixed one.
  - ✓ We show that the non-stationary MAP_2 is characterized by a set of five moments.
  - ✓ We define the unique canonical representation for the non-stationary MAP_2 in terms of five parameters (instead of six).
  - ✓ Every result presented has been illustrated with numerical examples.

- Motivated by the result of a canonical representation in the non-stationary MAP_2, we develop an inference procedure in Chapter 3.

  ✓ We introduce an important modeling framework based on N independent inter-event sequences, that can be extended to a wide range of applications.

  ✓ We provide the autocorrelation function of the inter-event times for the non-stationary version of the MAP_m.

  ✓ We develop a moment matching approach to model the sequence of failures of N electrical components to a non-stationary MAP_2.

  ✓ A real data case provided by the Spanish private electrical utility company, Iberdrola, has been provided to illustrate our developed approach. Important quantities regarding the counting process are calculated.

  ✓ We compare the capability of the non-stationary MAP_2 to properly capture the statistical characteristics of a sequence of failures of N electrical components versus its stationary counterpart.

- We have studied the stationary Batch Markovian arrival process with two-states, and batch arrivals up to size k, $BMAP_2(k)$ in Chapter 4.

  ✓ We have defined important distributional properties of the batch arrival process: the joint Laplace-Stieltjes transform between the inter-event times and batch sizes; and the autocorrelation function.

  ✓ We provide conditions under which two $BMAP_2(2)$s are equivalent for $n = 1$ and $n = 2$. And also give a procedure to construction an equivalent $BMAP_2(2)$ to a given one.

  ✓ We prove that the stationary $BMAP_2(2)$ is a nonidentifiable process.

  ✓ We have defined when a $BMAP_2(k)$, for $k \geq 3$, is nonidentifiable, based on the decomposition of nonidentifiable $BMAP_2(2)$s. We extend that result to prove that the stationary $BMAP_2(k)$, for $k \geq 3$, is a nonidentifiable process.

  ✓ Numerical examples have been presented to show the presented results.

- The autocorrelation functions of the inter-event times and batch sizes of the $BMAP_2(2)$ were examined in Chapter 5.

- ✓ We have characterized the autocorrelation function of the inter-event times in terms of the eigenvalue different from 1 of $P^{\star}$, for the stationary $BMAP_2(2)$. An analogous result was given for the autocorrelation function for the batch sizes.
- ✓ We prove that both correlation functions decrease geometrically as the time lag increases, for the stationary $BMAP_2(2)$.
- ✓ We identify four behavior patterns for both correlation functions. Numerical examples have been presented to show the presented results.
- ✓ We have extended the characterization of the autocorrelation functions of the inter-event times and batch sizes for the general stationary $BMAP_m(k)$.

Possible research lines as extensions of the presented work are the following,

- ↪ In Chapter 2 we studied the canonical representation of the non-stationary MAP_2. A possible extension of this work is to thoroughly study the identifiability problem for the non-stationary process when group or batch arrivals are allowed, $BMAP_m(k)$, for $k \geq 2$.

- ↪ In Chapter 3 we considered a framework of N electrical components identically built to fit a non-stationary MAP_2. A first extension is to consider a new setting where not all the components are equal. In such case, it would be desirable to develop an estimation technique that either groups or separates those components that are structurally the same. Second, it is of interest to derive an estimation approach when the repair times are non-negligible. Finally, we wish to study how the MAP can be estimated in the case of censored inter-failure times data, a very common problem in practice.

- ↪ In Chapter 4 we proved the non-identifiability of the stationary $BMAP_2(k)$, for $k \geq 2$. A challenging extension concerning the $BMAP$s is the study of the canonical representation of the process, a reduced/unique form that characterizes the process in terms of a fewer number of parameters (as Bodrog et al. [7] stated for stationary MAP_2).

  Other prospects regarding this chapter may concern both the estimation of the stationary $BMAP_2(k)$ and the study of the nonidentifiability of the higher order stationary $BMAP_m(k)$, for $m \geq 3$, which are expected to show more versatility for modeling purposes. Concerning the second point, we are aware of the complexity of such a problem due to the increasing number of parameters.

↪ In Chapter 5 we proved that the autocorrelation functions for the inter-event times and batch sizes decrease geometrically for the stationary $BMAP_2(2)$. An important extension would be to determine patterns and bounds for the autocorrelation functions for the stationary $BMAP_m(k)$, with $m \geq 3$. As well as study the joint correlation function of the inter-event times and batch sizes, and determine if greater structures can be captured.

↪ Extend the estimation methods proposed for the $BMAP$ to estimate the steady-state distributions of the queueing system. In addition, following Ausín et al. [5], we will also try to design a $BMAP/G/c$ queueing system, that is, given arrival and service data, our objective will be to choose the optimal number of servers so as to minimize an expected cost function which will depend on quantities, such as the number of customers in the queue.

↪ It is of interest to derive a hypothesis test given a sequence of inter-event times to determine which MAP adjusts better the data, i. e., try to find the MAP (with some order m, m the smallest as possible) which better fits the trace.

Bibliography

[1] Aalen O, Husebye E (1993). Statistical analysis of repeated events forming renewal processes. *Statistics in medicine*, 10, 1227-1240.

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

[3] Asmussen, S., Nerman O. and Olsson, M. (1996). Fitting phase type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23, 419-441.

[4] Asmussen, S.(2003).*Applied probability and queues*. Springer.

[5] Ausín, M., Lillo, R. and Wiper, M. (2007). Bayesian control of the number of servers in a $GI/M/c$ queueing system, *Journal of statistical planning and inference*, 137, 3043-3057.

[6] Blackwell, D. and Koopmans, L. (1957). On the identifiability problem for functions of finite Markov chains. *The Annals of Mathematical Statistics*, 28, 1011-1015.

[7] Bodrog, L., Heindl, 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.

[8] Bookbinder, J., Cai, Q. and He, QM. (2011). Shipment consolidation by private carrier: the discrete time and discrete quantity case. *Stochastic Models*, 27, 664-686.

[9] Breuer, L. (2002). An EM algorithm for Batch Markovian arrival processes and its comparison to a simpler estimation procedure, *Annals of Operations Research*, 121, 123-138.

[10] Breuer, L. and Baum, D. (2005). An introduction to queueing theory and matrix-analytic methods. Springer.

[11] Buchholz, P. (2003). An EM-algorithm for *MAP* fitting from real traffic data. In: Kemper, P., Sanders, W.H. (eds.) Computer Performance Evaluation/TOOLS. Lecture Notes in Computer Science, vol. 2794, pp. 218âĂŞ236. Springer, New York.

[12] Carrizosa, E. and Ramírez-Cobo, P. (2013). Maximum likelihood estimation for the two-state Markovian arrival process (submitted for publication) arXiv:1401.3105v1.

[13] Casale, G., Zhang, E. and Simirni, E. (2008). Interarrival times characterization and fitting for Markovian traffic analysis. Technical Report WM-CS-2008-02, available at http://www.wm.edu/as/computerscience/documents/cstechreports/WM-CS-2008-02.pdf, College of William and Mary.

[14] Casale, G., Zhang, E. and Simirni, E. (2010). Trace data characterization and fitting for Markov modeling. *Performance Evaluation*, 67, 61-79.

[15] Chakravarthy, SR. (2001). The batch Markovian arrival process: a review and future work. In: Krishnamoorthy A, et al., editors. Advances in probability theory and stochastic processes. NJ: Notable Publications, Inc. pp. 21-49.

[16] Chakravarthy, SR. (2010). The Markovian arrival processes: a review and future work.In J. Cochran, editor. Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons, Inc.

[17] Chakravarthy, SR. (2012). Maintenance of a deteriorating single server system with Markovian arrivals and random shocks. *European Journal of Operational Research*, 222, 508-522.

[18] Chakravarthy, SR and Neuts, MF. (2014). Analysis of a multi-server queueing model with *MAP* arrivals of regular customers and phase type arrivals of special customers. *Simulation Modelling Practice and Theory*, 34, 79-95.

[19] Coleman, T. and Li, Y. (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6, 418-445.

[20] Cordeiro, J. and Kharoufeh, J. (2010). Batch Markovian Arrival Processes (*BMAP*). Wiley Encyclopedia of Operations Research and Management Science.

[21] Cox, D.R. (1955). A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51, 313-319.

[22] Cumani, A. (1982). On the canonical representation of homogeneous Markov processes modeling failure-time distributions. *Microelectronics Reliability*, 22, 583âĂŞ602.

[23] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39, 1-38.

[24] Der Kiureghiana, A., Ditlevsenb, OD. and Song, J. (2007). Availability, reliability and downtime of systems with repairable components.*Reliability Engineering and System Safety*, 92, 231-422.

[25] Der Kiureghiana, A. and Song, J. (2008). Multi-scale reliability analysis and updating of complex systems by use of linear programming.*Reliability Engineering and System Safety*, 93, 288-297.

[26] Ephraim, Y. and Merhav, N. (2002). Hidden Markov processes. *IEEE Transactions on Information Theory*, 48, 1518-1569.

[27] 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.

[28] Falin, GI. (2010). A single-server batch arrival queue with returning customers. *European Journal of Operational Research*, 201, 786-790.

[29] Ferng, H-W. and J-F. Chang. (2001). Departure processes of BMAP/G/1 queues. *Queueing Systems*, 39, 109-135.

[30] Fearnhead, P. and Sherlock, C. (2006). An exact Gibbs sampler for the Markov-modulated Poisson process, *Journal of the Royal Statistical Society*, 68, 767-784.

[31] Finkelstein, M. and Marais, F. (2010) On terminating Poisson processes in some shock models. *Reliability Engineering and System Safety*, 93, 874-879

[32] Fiondella, L., Rajasekaran, S. and Gokhale SS. (2011). Efficient System Reliability with Correlated Component Failures. *IEEE 13th International Symposium on High-Assurance Systems Engineering (HASE)*, 269-276.

[33] Gillespie, J., McClean, S., Scotney, B., Garg, L., Barton, M. and Fullerton, K. (2011). Costing hospital resources for stroke patients using phase-type models. *Health Care Management Science*, 14, 279-291.

[34] Gómez-Corral, A. (2002). A tandem queue with blocking and Markovian arrival process. *Queueing Systems*, 41, 343-370.

[35] Gómez-Corral, A. and Economou, A. (2007). The batch Markovian arrival process subject to renewal generated geometric catastrophes. *Stochastic Models*, 23, 211-233.

[36] He, QM. (2014). *Fundamentals of Matrix-Analytic Methods*. Springer.

[37] He, QM. and Zhang H. (2006). PH-invariant polytopes and coxian representations of phase type distributions. *Stochastic Models*, 22, 383-409.

[38] He, QM. and Zhang H. (2008). An algorithm for computing minimal coxian representations. *INFORMS Journal on Computing*, 20, 179-190.

[39] He, QM. and Zhang H. (2009). Coxian representations of generalized Erlang distributions. *Acta Mathematicae Applicatae Sinica*, 25, 489-502.

[40] Heckmüller, S. and Wolfinger, BE. (2008). Using load transformations to predict the impact of packet fragmentation and losses on Markovian arrival processes: In: K. Al-Begain, A. Heindl & M. Telek (eds.), Proceedings of ASMTA 2008. Springer, Berlin, pp 31-46.

[41] Heindl, A., Mitchell, K. and van de Liefvoort, A. (2006). Correlation bounds for second-order *MAP*s with application to queueing network decomposition. *Performance Evaluation*, 63, 553-577.

[42] Hervé, L. and Ledoux, J. (2013). Geometric ρ-mixing property of the interarrival times of a stationary Markovian arrival process. *Journal of Applied Probability*, 50, 598-601.

[43] Horváth, G. and Telek, M. (2002). Markovian modeling of real data traffic: Heuristic phase type and *MAP* fitting of heavy tailed and fractal like samples. In: Calzarossa, M.C., Tucci, S. (eds.) Proceedings of the Performance 2002. Lecture Notes in Computer Science, vol. 2459,pp. 405-434. Springer, Berlin.

[44] Horváth, G. and Telek, M. (2009). On the canonical representation of phase type distributions. *Performance Evaluation*, 66, 396-409.

[45] IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. IEEE Press Piscataway, New York, NY; 1990.

[46] Ito, H., Armari, S-I. and Kobayashi, K. (1992). Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Transactions on Information Theory*, 38, 324-333.

[47] Kang, S., Han Kim, Y., Sung, D. and Choi, B. (2002). An application of Markovian arrival process to modeling superposed ATM cell streams. *IEEE Transactions on Communications*, 50, 633-642.

[48] Kim, CS., Klimenok, VI. and Orlovskii, DS. (2008). The $BMAP/PH/N$ retrial queue with Markovian flow of breakdowns. *European Journal of Operational Research*, 189, 1057-1072.

[49] Kim, B. and Kim, J. (2010). Queue size distribution in a discrete time $D-BMAP/G/1$ retrial queue. *Computers & Operations Research*, 37, 1220-1227.

[50] Kim, CS., Klimenok, VI., Mushko, V. and Dudin, A. (2010). The $BMAP/PH/N$ retrial queueing system operating in Markovian random environment. *Computers & Operations Research*, 37, 1228-1237.

[51] Klemm, A., Lindemann, C. and Lohmann, M. (2003). Modeling IP traffic using the batch Markovian arrival process. *Performance Evaluation*, 54, 149-173.

[52] Kriege, J. and Buchholz, P. (2010). An empirical comparison of *MAP* fitting algorithms. In B. Muller-Clostermann, K. Echtle, and E. P. Rathgeb (eds.), Lecture Notes in Comput. Sci., vol. 5987. Springer, Berlin, pp 259-73.

[53] Landriault, D. and Shi, T. (2014). Occupation times in the *MAP* risk model. *Insurance: Mathematics and Economics*,doi = http://dx.doi.org/10.1016/j.insmatheco.2014.10.014.

[54] Latouche, G. and Ramaswami, V. (1990). Introduction to Matrix Analytic Methods in Stochastic Modeling, ASA-SIAM Series on Statistics and Applied Probability, vol. 5, SIAM, Philadelphia, PA.

[55] Lawler, G. (2006). Introduction to Stochastic Processes. Second Edition.Chapman & Hall/CRC Probability Series.

[56] Leroux, B. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications*, 40, 127-143.

[57] Levitin, G., Xing, L. and Dai, Y. (2014). Optimal component loading in 1-out-of-N cold standby systems. *Reliability Engineering and System Safety*, 127, 58-64.

[58] Li, Q-L. and Zhao, Y.Q. (2004). A $MAP/G/1$ queue with negative customers. *Queueing Systems*, 47, 4-43.

[59] Li, Q-L.,Ying, Y. and Zhao, Y.Q. (2006). A $BMAP/G/1$ retrial queue with a server subject to breakdowns and repairs. *Annals of Operations Research*, 141, 233-270.

[60] Li, S. and Ren, J. (2013). The maximum severity of ruin in a perturbed risk process with Markovian arrivals *Statistics & Probability Letters*, 83, 993 -998.

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

[62] Lucantoni, D. (1991). New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46.

[63] Lucantoni, D. (1993). The $BMAP/G/1$ queue: a tutorial. In: L. Donatiello and R. Nelson (eds.) Models and Techniques for performance Evaluation of Computer and Communications Systems. Springer Verlag, Berlin, pp 330-358.

[64] MacCluer, C. R.(2000). The Many Proofs and Applications of PerronâĂŹs Theorem. *SIAM Review*, 3, 487-498.

[65] Mahé, C. and Chevret, S. (2001). Analysis of recurrent failure times data: should the baseline hazard be stratified?. *Statistics in Medicine*, 20, 3807-3815.

[66] Marshall, A., Mitchell, H. and Zenga, M. (2014). Modelling the Length of Stay of Geriatric Patients in the Emilia Romagna Hospitals Using Coxian Phase-Type Distributions with Covariates. *Studies in Theoretical and Applied Statistics*, doi=10.1007/10104_2014_21.

[67] Montoro-Cazorla, D. and Pérez-Ocón, R. (2006). Reliability of a system under two types of failures using Markovian arrival processes. *Operations Research Letters*, 34, 525-530.

[68] Montoro-Cazorla, D., Pérez-Ocón R. and Segovia MC. (2007) Survival probabilities for shock and wear models governed by phase-type distributions. *Quality Technology & Quantitative Management*, 4, 85-94.

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

[70] Montoro-Cazorla, D. and Pérez-Ocón, R. (2012) A shock and wear system under environmental conditions subject to internal failures, repair, and replacement. *Reliability Engineering and System Safety*, 99, 55-61.

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

[72] Montoro-Cazorla, D. and Pérez-Ocón, R. (2014). A redundant n-system under shocks and repairs following Markovian arrival processes. *Reliability Engineering and System Safety*, 130, 69-75.

[73] Nakagawa, T. (2011). *Stochastic Processes with Applications to Reliability Theory.* Springer Series in Reliability Engineering.

[74] Neuts, MF. (1975). Probability distributions of phase type. In: R. Holvoet and H. Florin (eds.) Liber Amicorum Prof. Emeritus H. Florin, University of Louvain. Belgium, pp 173-206.

[75] Neuts, MF. (1979). A versatile Markovian point process. *Journal of Applied Probability*, 16, 764-779.

[76] Neuts, MF., Bhattacharjee MC. (1981). Shock models with phase type survival and shock resistance. *Naval Research Logistic* 28, 213-219.

[77] Neuts, MF., and Li, JM. (1996). An algorithm for the $P(n, t)$ matrices of a continuous *BMAP*. In: Chakravarthy SR, Alfa, AS, editors. Matrix-analytic methods in stochastic models. NY: Marcel Dekker; 7-19.

[78] Niyato, D., Hossain, E. and Fallahi, A. (2007). Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: performance analysis and optimization. *IEEE Transactions on Mobile Computing*, 6, 221-236.

[79] O´ Cinneide, C. (1990). On nonuniqueness of representations of phase-type distributions.Communications in Statistics. *Stochastic Models*, 5, 247-259.

[80] Okamura, H., Dohi, T. and Trivedi, K. (2009). Markovian arrival process parameter estimation with group data. *IEEE/ACM Trans. Networking*, 17, 1326-339.

[81] Oliver, I. (2009). *Markov Processes for Stochastic Modeling.* First Edition. Elsevier Academic Press.

[82] Peng, D., Fang, L. and Tong, C. A multi-state reliability analysis of single-component repairable system based on phase-type distribution. 2013 International Conference in Management Science and Engineering (ICMSE). IEEE 2013: 496-501.

[83] Pérez-Ocón, R. and Montoro-Cazorla, D. (2004). A multiple system governed by a quasi-birth-and-death process. *Reliability Engineering and System Safety*, 84, 187-196.

[84] Pérez-Ocón, R. and Montoro-Cazorla, D. A shock and wear system under preventive and corrective repairs. Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013. 2014: 3135-3141.

[85] Proschan, F. (1963). Theoretical Explanation of observed decreasing Failure Rate. *Technometrics*, 5, 375-383.

[86] Ramaswami, V. (1980). The N/G/1 queue and it detailed analysis. *Advanced Applied Probability*, 12, 222-261.

[87] Ramírez-Cobo, P., Lillo, R.E. and Wiper, M. (2010). Non identifiability of the two-state Markovian arrival process. *Journal of Applied Probability*, 47, 630-649.

[88] Ramírez-Cobo, P., Lillo, R.E. and Wiper, M. (2014). Identifiability of the $MAP_2/G/1$ queueing system. *TOP* , 22, 274-289.

[89] Ramírez-Cobo, P, Marzo, X., Olivares-Nadal, A., Álvarez Francoso, J. and Carrizosa, E. (2014). The Markovian arrival process: A statistical model for daily precipitation amounts. *Journal of Hydrology*, 510, 459-471.

[90] P. Ramírez-Cobo, R. Lillo, and M. Wiper. (2014). Bayesian inference for the two-state Markovian arrival process. *Submitted*.

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

[92] 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.

[93] Rydén T. (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *Annals of Statistics*, 22, 1884-1895.

[94] Rydén, T. (1996). On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions. *Journal of Applied Probability*, 33, 640-653.

[95] Rydén, T. (1996). An EM algorithm for estimation the Markov Modulated Poisson processes, *Computational Statistics and Data Analysis*, 21, 431-447.

[96] Rodríguez L. Maximum likelihood estimation of phase-type distributions. Technical University of Denmark (DTU) 2011.

[97] Rodríguez, J., Lillo, R.E. and Ramírez-Cobo, P. (2014). Nonidentifiability of the two-state *BMAP*, *Methodology and Computing in Applied Probability.* http://dx.doi.org/10.1007/s11009-014-9401-z.

[98] Rodríguez, J., Lillo, R.E. 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.

[99] Ross, S. (2009). *A First Course in Probability.* Eighth Edition. Pearson Prentice Hall.

[100] Scott, S. (1999). Bayesian analysis of the two state Markov Modulated Poisson process. *Journal of Computational and Graphical Statistics*, 8, 662-670.

[101] Song, S., Coit, D., Feng, Q. and Peng, H. (2014). Reliability analysis for multi-component systems subject to multiple dependent competing failure processes. *IEEE Transactions on Reliability*, 63, 331-345.

[102] Telek, M. and Horváth, G. (2007). A minimal representation of Markov arrival processes and a moments matching method. *Performance Evaluation*, 64, 1153-1168.

[103] Tseng, SM. and Wang, YC. (2013). Throughput of DS-CDMA/unslotted ALOHA radio networks with Markovian arrival processes. *International Journal of Communication Systems*, 26, 369-379.

[104] Vilar, J.M., Cao, R., Ausín M.C. and González-Fragueiro, C. (2009). Nonparametric analysis of aggregate loss models. *Journal of Applied Statistics*, 36, 149-166.

[105] Wasserman, L. (2004) All of Statistics: A Concise Course in Statistical Inference. Springer Texts in Statistics.

[106] 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 Modeling*, 35, 1184-193.