ISSN: 1696-2281
eISSN: 2013-8830
SORT 34 (2) July-December (2010)

SORT

Statistics and Operations Research Transactions

Sponsoring institutions

Universitat Politècnica de Catalunya
Universitat de Barcelona
Universitat de Girona
Universitat Autònoma de Barcelona
Institut d'Estadística de Catalunya

Supporting institution

Spanish Region of the International Biometric Society

Generalitat de Catalunya
Institut d'Estadística de Catalunya


FSC
www.fsc.org
100%
Paper procedent de
boscos ben gestionats
FSC™ C081765

SORT

Volume 34
Number 2
July-December 2010
ISSN: 1696-2281
eISSN: 2013-8830

Invited article *(with discussion)*

Articles

Selected article from *XII Conferencia Española de Biometría 2009*

Book review

Information for authors and subscribers

SORT 34 (2) July-December 2010, 101-144

Markovian arrivals in stochastic modelling: a survey and some new results

Jesús R. Artalejo, Antonio Gómez-Corral

Faculty of Mathematics, Complutense University of Madrid, Madrid 28040, Spain

Qi-Ming He

Department of Management Sciences, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Abstract

This paper aims to provide a comprehensive review on *Markovian arrival processes* (MAPs), which constitute a rich class of point processes used extensively in stochastic modelling. Our starting point is the versatile process introduced by Neuts (1979) which, under some simplified notation, was coined as the *batch Markovian arrival process* (BMAP). On the one hand, a general point process can be approximated by appropriate MAPs and, on the other hand, the MAPs provide a versatile, yet tractable option for modelling a bursty flow by preserving the Markovian formalism. While a number of well-known arrival processes are subsumed under a BMAP as special cases, the literature also shows generalizations to model arrival streams with marks, non-homogeneous settings or even spatial arrivals. We survey on the main aspects of the BMAP, discuss on some of its variants and generalizations, and give a few new results in the context of a recent state-dependent extension.

MSC: 60Jxx, 60G55

Keywords: Markovian arrival process, batch arrivals, marked process, phase-type distribution, BSDE approach

1. Introduction

The *versatile Markovian point process* introduced by Neuts (1979) was the seminal work, in conjunction with the *phase* (PH) type distribution, for getting beyond two common and extended assumptions in stochastic modelling, namely: (a) the exponential

Received: April 2010

distribution and the *Poisson process* (PP), which are the key tools for constructing Markovian models; and (b) the independence and equidistribution of the successive inter-arrival intervals, which are inherent features of the PP and the renewal processes. Later, it was proved that the *batch Markovian arrival process* (BMAP) is equivalent to the versatile Neuts process. Since the former presents a more transparent notation, at present it is widely accepted to refer to the BMAP rather than to the Neuts process.

The popularity of the BMAP and other Markovian arrival processes comes from the following important features:

(i) They provide a natural generalization of the PP and the renewal processes.
(ii) They take into account the correlation aspect, which arises naturally in many applications where the arrival flow is bursty.
(iii) They preserve the tractable Markovian structure.

As a result, the use of Markovian arrival processes in combination with the impetus provided by the modern computational advances explains the spectacular growth of applications to queueing, inventory, reliability, manufacturing, communication systems, and risk and insurance problems.

The use of BMAPs and PH distributions in stochastic modelling readily leads to the so called matrix-analytic formalism where scalar quantities are replaced by matrices. The main resulting structured Markov chains have been extensively studied; see the monographs by Bini *et al.* (2005), Latouche and Ramaswami (1999), Li (2010) and Neuts (1981,1989). Qualitatively, the consideration of the BMAP for modelling the arrival input greatly enhances the versatility of the stochastic model. For practical use, presenting the model under a suitable structured matrix form makes it easy to be studied in a unified manner and in an algorithmically tractable way. However, it should be pointed out that the cost lies in the risk of finding computational problems derived from an excessive dimensionality caused by the matrix formalism.

This survey paper is aimed on providing information on Markovian arrival processes, putting emphasis on the discussion of extensions and variants of the BMAP, as well as on the wide use of this class of processes in applications. Following the leads in this paper and the guidance provided by the bibliographical notes, readers can get access to the background materials where technical details and proofs are available.

This survey is organized as follows. In Section 2, we first introduce the BMAP and the continuous PH distribution. A number of important particular cases, the basic properties and descriptors of the BMAP, as well as some applications in queueing, reliability and inventory models are presented in subsequent sections. In Section 3, we consider a number of generalizations and variants of the BMAP including the discrete counterpart (D-BMAP), the *marked Markovian arrival process* (MMAP), the *HetSigma* approach, the Markov-additive processes of arrivals and the *block-structured state-dependent event* (BSDE) approach. The consideration of these extensions and variants enriches the methodology and enhances the versatility of the arrival processes

in different directions. Based on the fact that the BSDE approach allows us to deal with modulated non-homogeneous settings, but keeping the dimensionality of the underlying matrices tractable, Section 4 applies this approach to the SIS epidemic model. Some new results concerning with the extinction time and the correlation between events are obtained. We conclude the survey with a few bibliographical notes. A glossary of notation is presented in Appendix.

2. The BMAP

The PP is the basic renewal process where inter-renewal times are exponentially distributed. The PH distribution and the BMAP can be thought of as the natural generalizations of the exponential distribution and the PP, respectively. They are both based on the method of stages, which was introduced by A.K. Erlang and extensively generalized by M.F. Neuts. On the other hand, the PH distribution and the BMAP can be viewed as particular cases of the matrix-exponential distribution and the rational arrival process; the interested reader is referred to the papers by Asmussen and Bladt (1999), and Nielsen *et al.* (2007).

Although our main interest is put on the BMAP and its extensions, the PH distribution is used many times along the paper. Thus, before focussing on a description of the BMAP, we briefly introduce the continuous PH distribution.

The class of probability distributions of PH type provides a simple framework to demonstrate how one may extend many results on exponential distributions to more complex models, but without losing computational tractability. The key idea is to exploit the fact that many distributions derived from the exponential law can be formulated as the distribution of the time till absorption in suitably defined Markov processes. This allows one to deal with PH distributions by appealing to the simple dependence structure underlying Markov processes.

To define a PH distribution we consider an absorbing Markov chain on the state space $\{0,1,\ldots,n\}$ with initial probability vector $(1-\boldsymbol{\tau}\mathbf{e}_n,\boldsymbol{\tau})$ and infinitesimal generator

$$\begin{pmatrix} 0 & \mathbf{0}_n \\ \mathbf{t} & \mathbf{T} \end{pmatrix},$$

where $\mathbf{t}=-\mathbf{T}\mathbf{e}_n$. Then, a PH distribution corresponds to the distribution of the time L until absorption into the state 0. Thus, we have the following expressions for the distribution function, the density function and the moments:

$$\begin{aligned} F(x) &= 1-\boldsymbol{\tau}\exp\{\mathbf{T}x\}\mathbf{e}_n, \quad x\geq 0,\\ f(x) &= \boldsymbol{\tau}\exp\{\mathbf{T}x\}\mathbf{t}, \quad x\geq 0,\\ E\left[L^k\right] &= k!\boldsymbol{\tau}\left(-\mathbf{T}^{-1}\right)^k\mathbf{e}_n, \quad k\geq 1. \end{aligned}$$

An important question to be examined is when the absorption occurs in a finite interval almost surely. By using the above expression for the distribution function, it is readily verified that $F(\infty) = 1$ if and only if the matrix $\mathbf{T}$ is non-singular. Furthermore, this is certain if and only if states in $\{1,\ldots,n\}$ are all transient.

For practical use, the class of PH distributions provides ease in conditioning arguments, results in a Markovian structure of models involving exponential assumptions and leads to significant simplifications in various integral and differential equations arising in their analysis. An excellent summary of closure properties can be found in Asmussen (2000), Latouche and Ramaswami (1999, Section 2.6) and Neuts (1981, Chapter 2). Among these, we emphasize three properties. First, this class is dense, in the sense of weak convergence, in the class of all distributions on $[0,\infty)$. Second, sums and mixtures of a finite number of independent PH random variables are PH random variables. Third, all order statistics of a set of independent PH random variables are themselves PH random variables.

The PP has served as the main arrival flow for many years and generalizations have frequently concentrated on renewal processes. Their simplifying feature is the independence and equidistribution of successive inter-renewal intervals. Thus, in queueing and other applications (see Neuts (1992)), the class of renewal processes is not flexible enough and, in particular, arrivals that tend to occur in bursts cannot be modelled in this way.

We present here the BMAP, which is thought to be a fairly general point process where the correlation aspect is not ignored. It is, in general, a non-renewal process having the feature of making many analytic properties explicit or at least computationally tractable. The key idea is to generate counting processes by modelling the transitions of a Markov chain; see also Rudemo (1973).

We begin with a constructive description of the BMAP. The BMAP is a bivariate Markov process $\{(N(t),J(t));t \geq 0\}$ on $\mathscr{S} = \mathbb{N} \times \{1,\ldots,m\}$, where $N(t)$ represents the number of arrivals up to time t, while the states of the background Markov chain $\{J(t);t \geq 0\}$ are called phases. Let us assume that $m < \infty$ and denote by $\mathbf{D}$ the infinitesimal generator of the background Markov chain, which is assumed to be irreducible. At the end of a sojourn time in $(n,i) \in \mathscr{S}$, which is exponentially distributed with parameter λ_i, there occurs a transition to another or (possibly) the same phase state. That transition may or not correspond to an arrival epoch. Specifically, with probability $P_{ij}(k)$, it corresponds to a transition to state j with a batch arrival of size k, for $k \geq 1$, and similarly, with probability $P_{ij}(0)$, the transition corresponds to no arrival and state of the underlying Markov chain is j, for $j \neq i$. Therefore, $J(t)$ can go from state i to state i only through an arrival and

$$\sum_{j=1,\ j\neq i}^{m} P_{ij}(0) + \sum_{j=1}^{m}\sum_{k=1}^{\infty} P_{ij}(k) = 1, \quad 1 \leq i \leq m.$$

Define the matrices $\mathbf{D}_k = (d_{ij}(k))$ with entries $d_{ii}(0) = -\lambda_i$, $d_{ij}(0) = \lambda_i P_{ij}(0)$, for $j \neq i$, and $d_{ij}(k) = \lambda_i P_{ij}(k)$, for $k \geq 1$, from which it is clear that $\mathbf{D} = \sum_{k=0}^{\infty} \mathbf{D}_k$. The particular

choice $\mathbf{D}_0 \neq \mathbf{D}$ and $\mathbf{D}_k = \mathbf{0}_{m\times m}$, for $k \geq 2$, means single arrivals and yields the *Markovian arrival process* (MAP). In this formulation, the introduction of phases is the key to get dependent non-exponential inter-arrival time distributions, and correlated batch sizes.

Our preceding construction shows that the bivariate process $\{(N(t), J(t)); t \geq 0\}$ has the structured infinitesimal generator

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \cdots \\ & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \cdots \\ & & \mathbf{D}_0 & \mathbf{D}_1 & \cdots \\ & & & \ddots & \ddots \end{pmatrix}.$$

The sequence of matrices $\{\mathbf{D}_k; k \geq 0\}$ contains all information for $\mathbf{Q}$ and thus is usually called the characteristic sequence of a BMAP. Although we often ignore the determination of $J(0)$, a complete specification requires specification of the distribution of $J(0)$. We may do this in terms of a row vector $\boldsymbol{\alpha}$ with ith entry given by $P(J(0) = i)$, for $1 \leq i \leq m$.

By assuming $\mathbf{D}_0$ to be non-singular, the inter-arrival times are finite, with probability one. An additional assumption is that the vector $\mathbf{d} = \overline{\mathbf{D}}_1 \mathbf{e}_m$ is finite, where $\overline{\mathbf{D}}_1 = \sum_{k=1}^{\infty} k\mathbf{D}_k$. This condition is equivalent to require that $E[N(t)] < \infty$ over finite intervals. The fundamental arrival rate is then defined by $\lambda = \boldsymbol{\theta}\mathbf{d}$, where $\boldsymbol{\theta}$ is the unique positive probability vector satisfying $\boldsymbol{\theta}\mathbf{D} = \mathbf{0}_m$ and $\boldsymbol{\theta}\mathbf{e}_m = 1$, and consequently it amounts to the expected number of single arrivals per unit of time in the stationary version of a BMAP.

This family of counting processes has received several names in the literature. The currently used term batch Markovian arrival process evolved from versatile Markovian point process (see Neuts (1979)) and *Neuts process* (see Ramaswami (1980)) to *non-renewal arrival process* (see Lucantoni *et al.* (1990)), until it was settled down at batch Markovian arrival process by Lucantoni (1991). Lucantoni (1991) also introduced a simple matrix representation for the BMAP, which made it easy to interpret parameters of Markovian arrivals and to use this class of arrival processes in stochastic modelling.

We next present two alternative definitions of the BMAP and a few examples of BMAPs with special characteristics.

Remark 2.1 The BMAP can be thought of as a semi-Markovian arrival process. Define the sequence $\{(J_n, K_n, \tau_n); n \geq 0\}$, where J_n is the phase of $\{J(t); t \geq 0\}$ right after the nth batch arrival, K_n is the size of the nth batch, and τ_n is the inter-arrival time between the $(n-1)$st and the nth arrival events. Then, $\{(J_n, K_n, \tau_n); n \geq 0\}$ satisfies

$$\begin{aligned} P(J_n = j, K_n = k, \tau_n \leq x | J_{n-1} = i) &= \left(\int_0^x \exp\{\mathbf{D}_0 u\} du \mathbf{D}_k \right)_{ij} \\ &= \left((\mathbf{I}_m - \exp\{\mathbf{D}_0 x\}) \left(-\mathbf{D}_0^{-1}\right) \mathbf{D}_k \right)_{ij}, \end{aligned}$$

for $1 \leq i, j \leq m$, $k \geq 1$ and $x \geq 0$.

Remark 2.2 Equivalently, we may present a definition of the BMAP based on PPs. Let m be a finite positive integer, $\{\alpha_i; 1 \le i \le m\}$ be non-negative numbers satisfying $\sum_{i=1}^{m} \alpha_i = 1$, and $\{d_{ij}(0); 1 \le i, j \le m, j \ne i\}$ and $\{d_{ij}(k); 1 \le i, j \le m\}$, for $k \ge 1$, be non-negative numbers. Assume that $-d_{ii}(0) > 0$, where

$$-d_{ii}(0) = \sum_{j=1,\ j\ne i}^{m} d_{ij}(0) + \sum_{j=1}^{m}\sum_{k=1}^{\infty} d_{ij}(k), \quad 1 \le i \le m.$$

The bivariate process $\{(N(t), J(t)); t \ge 0\}$ can be defined as follows:

(i) Define independent PPs with parameters $d_{ij}(0)$, for $1 \le i, j \le m$ and $j \ne i$, and $d_{ij}(k)$, for $1 \le i, j \le m$ and $k \ge 1$. If $d_{ij}(k) = 0$, then the corresponding PP has no event.
(ii) Determine $J(0)$ by the probability distribution $\{\alpha_i; 1 \le i \le m\}$. Set $N(0) = 0$.
(iii) If $J(t) = i$, for $1 \le i \le m$, we let $J(t)$ and $N(t)$ remain the same until the first event occurs in the set of PPs with rates $d_{ij}(0)$, for $1 \le i, j \le m$ and $j \ne i$, and $d_{ij}(k)$, for $1 \le i, j \le m$ and $k \ge 1$. If the next event comes from the PP of rate $d_{ij}(0)$, then $J(t)$ changes from phase i to phase j and $N(t)$ does not change at this epoch, for $1 \le j \le m$ and $j \ne i$. On the contrary, if the next event comes from the PP of rate $d_{ij}(k)$, then the phase variable $J(t)$ transits from phase i to phase j, and $N(t)$ is increased by k units at this epoch, for $1 \le j \le m$ and $k \ge 1$; in this case, a batch of k units is associated with the event.

For use in simulations, it is easy to generate realizations of a BMAP from the dynamics described in Remark 2.2. The visualization of simulated paths of a BMAP, and their effect as input streams to queues, is an excellent way for practitioners to appreciate the versatility of this class of point processes; see Figure 1 in Example 2.1.

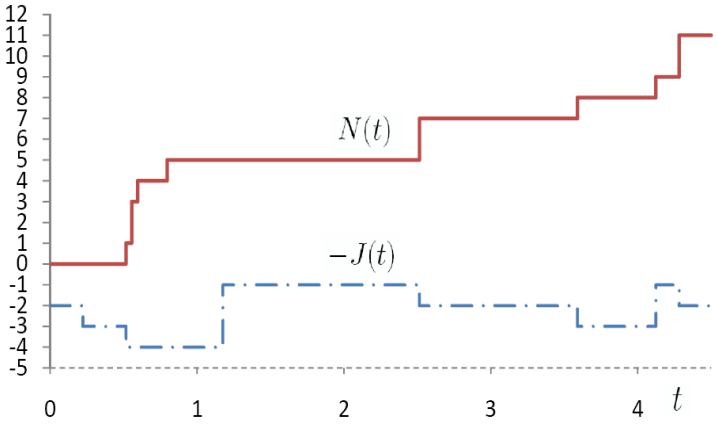


Figure 1: *A simulated sample path of a BMAP.*

Example 2.1 Consider a BMAP with non-null characteristic matrices

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 1 & 0 & -5 & 0 \\ 2 & 0 & 0 & -10 \end{pmatrix}, \ \mathbf{D}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \ \mathbf{D}_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Figure 1 shows a typical sample path of the bivariate process $\{(N(t), J(t)); t \geq 0\}$.

The following three choices of the BMAP are related to special characteristics:

(i) Bursty arrivals

$$\mathbf{D}_0 = \begin{pmatrix} -50 & 0 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{D}_1 = \begin{pmatrix} 49 & 1 \\ 0 & 0 \end{pmatrix}.$$

A widely accepted definition of burstiness does not exist; instead, several different measures can be used. In this paper, we assume the definition given by Neuts (1993). Qualitatively, the process is bursty as, over intervals of significant length, the actual number of arrivals is far in excess or far below the average. Positive autocorrelation between inter-arrival times explains, to a large extend, traffic burstiness. Obviously, the PP has independent inter-arrival times so it is not the appropriate model in case of bursty traffic.

(ii) Cyclic arrivals

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{D}_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}.$$

In this case, batches of size 1 and batches of size 2 arrive cyclically.

(iii) Bursty vs smooth

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 0 \\ 0 & -50 \end{pmatrix}, \quad \mathbf{D}_1 = \begin{pmatrix} 0 & 0 \\ 1 & 49 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The process related to batches of size 1 is bursty, while for batches of size 2 the process is smooth.

In Subsection 2.1, we give a few examples to illustrate the variety of models subsumed under the matrix formulation of a BMAP as special cases. Subsection 2.2 begins by introducing the time-dependent distribution of the bivariate process $\{(N(t), J(t)); t \geq 0\}$. We then examine basic properties that make the BMAP a versatile class for modelling purposes. We present in Subsection 2.3 some interesting descriptors. Our focus in Subsection 2.4 is on four examples showing the interest of the BMAP in different applications, such as reliabilty, queueing and inventory problems.

2.1. Particular cases

We describe in this subsection several special cases of the BMAP. We begin by listing a selected sample of processes obtained as particular cases of the MAP.

(i) *Poisson process*. The PP of rate $\lambda > 0$ corresponds to the simple scalar case where $m = 1$, $\mathbf{D}_0 = -\lambda$ and $\mathbf{D}_1 = \lambda$.

(ii) *Markov modulated Poisson process* (MMPP). The MMPP is a PP whose rate varies according to a finite Markov chain serving as a random environment. Let $\mathbf{Q}_a$ be its underlying infinitesimal generator. The arrival rate is $\delta_i > 0$ when the random environmental state is i. Then, the MMPP is a MAP with $\mathbf{D}_0 = \mathbf{Q}_a - \mathbf{\Lambda}$ and $\mathbf{D}_1 = \mathbf{\Lambda}$, where $\mathbf{\Lambda} = \text{diag}(\delta_1, \ldots, \delta_m)$.

(iii) *PH renewal process*. This is a renewal process in which the inter-renewal times follow a PH distribution with representation $(\boldsymbol{\tau}, \mathbf{T})$. Thus, we have the correspondence $\mathbf{D}_0 = \mathbf{T}$ and $\mathbf{D}_1 = \mathbf{t}\boldsymbol{\tau}$.

(iv) *A sequence of PH inter-arrival times governed via a Markov chain*. This process is also named *PH semi-Markov process*; see Latouche and Ramaswami (1999). Consider l PH distributions with representations $(\boldsymbol{\tau}_i, \mathbf{T}_i)$ of order n_i, for $1 \leq i \leq l$ and $\sum_{i=1}^{l} n_i = m$. The successive inter-arrival distributions are selected from these PH distributions according to a discrete Markov chain with one-step transition probability matrix $\mathbf{P}_a = (p_{ii'})$ of dimension l. We then have $\mathbf{D}_0 = \text{diag}(\mathbf{T}_1, \ldots, \mathbf{T}_l)$ and $\mathbf{D}_1 = (d_{ii'}(1))$, where $d_{ii'}(1) = t_i p_{ii'} \tau_{i'}$, for $1 \leq i, i' \leq l$, with $\mathbf{t} = (t_i)$ and $\boldsymbol{\tau} = (\tau_i)$. The choice $l = 2$ and $p_{12} = p_{21} = 1$ leads to an *alternating PH renewal process*.

It should be noted that the PH renewal process can be viewed as the trivial special case of (iv), where all the PH distributions are chosen to be identical. More interesting is the *Markov switched Poisson process* (MSPP) obtained by choosing the PH distributions as exponential distributions of rate $\delta_i > 0$; see Chakravarthy (2001). We also remark that the modulation in the MSPP is of a discrete nature and it occurs at arrival epochs, whereas the modulation of the MMPP is performed in continuous time.

We now give some examples where arrivals occur properly in batches.

(v) *Compound Poisson process* (CPP). The classical scalar PP with batch arrivals of rate $\lambda > 0$ and jump size distribution $\{g_k; k \geq 1\}$ is a BMAP with $m = 1$, $\mathbf{D}_0 = -\lambda$ and $\mathbf{D}_k = \lambda g_k$, for $k \geq 1$.

(vi) *MAP with i.i.d. batch arrivals*. A MAP with independent and identically distributed batch arrivals amounts to a BMAP with $\mathbf{D}_0 = \mathbf{D}_0^a$ and $\mathbf{D}_k = g_k \mathbf{D}_1^a$, for $k \geq 1$, where the pair $(\mathbf{D}_0^a, \mathbf{D}_1^a)$ is the representation of the underlying MAP of order m. This example shows a choice of the BMAP where the batch size does not depend on phase transitions.

(vii) *Bacth PH semi-Markov process*. This process is the batch version of (iv) in which $d_{ii'}(k) = g_k t_i p_{ii'} \tau_{i'}$, for $k \geq 1$. A *batch Markov switched Poisson process* (BMSPP) follows by reducing the PH distribution to the exponential case.

(viii) *Batch PP with correlated batch arrivals*. This is a CPP where the jump size distribution is selected according to a Markov chain with one-step transition probability matrix $\mathbf{P}_a$ of dimension m. The resulting BMAP has matrices $\mathbf{D}_0 = -\lambda \mathbf{I}_m$ and $\mathbf{D}_k = (d_{ii'}(k))$, where $d_{ii'}(k) = \lambda g_{ik} p_{ii'}$, for $1 \leq i, i' \leq m$ and $k \geq 1$. The notation g_{ik} stands for the probability that a batch of size k arrives when the phase state is i.

We notice that the auxiliary transition matrix is used in the MSPP to modulate arrival rates. However, the role of $\mathbf{P}_a$ in the batch PP with correlated arrivals is to modulate jump sizes.

2.2. Basic properties of the BMAP

We are next interested in the counting component $N(t)$ of the BMAP, the superposition and thinning mechanisms, the local poissonification of a MAP and the denseness property.

2.2.1. The counting function

Consider the matrices $\mathbf{P}(n,t)$, for $n \geq 0$ and $t \geq 0$, with (i,j)th element

$$P_{ij}(n,t) = P\left(N(t) = n, J(t) = j | N(0) = 0, J(0) = i\right), \quad 1 \leq i, j \leq m.$$

From the Kolmogorov forward equations of the process $\{(N(t), J(t)); t \geq 0\}$, we obtain

$$\frac{d\mathbf{P}(n,t)}{dt} = \sum_{k=0}^{n} \mathbf{P}(k,t)\mathbf{D}_{n-k}, \quad n \geq 1, t \geq 0,$$

and the initial condition $\mathbf{P}(0,0) = \mathbf{I}_m$.

The corresponding matrix generating function $\mathbf{P}^*(z,t) = \sum_{n=0}^{\infty} z^n \mathbf{P}(n,t)$, for $|z| \leq 1$ and $t \geq 0$, is given by the exponential matrix

$$\mathbf{P}^*(z,t) = \exp\{\mathbf{D}^*(z)t\},$$

with $\mathbf{D}^*(z) = \sum_{k=0}^{\infty} z^k \mathbf{D}_k$, for $|z| \leq 1$. The numerical computation of $\mathbf{P}(n,t)$ can be based on the uniformization method; see Neuts and Li (1997).

By routine calculations, we can find that the first moment matrix $\mathbf{M}_1(t)$ and the column vector $\mathbf{M}_1(t)\mathbf{e}_m$ are given by

$$\mathbf{M}_1(t) = \left.\frac{\partial \mathbf{P}^*(z,t)}{\partial z}\right|_{z=1} = \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{k=0}^{n-1} \mathbf{D}^k \overline{\mathbf{D}}_1 \mathbf{D}^{n-1-k},$$

$$\mathbf{M}_1(t)\mathbf{e}_m = \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{D}^{n-1} \overline{\mathbf{D}}_1 \mathbf{e}_m.$$

By using the above expression, it can be shown (see Neuts (1989)) that the Palm function $E[N(t)]$ is given by

$$E[N(t)] = \lambda t + \boldsymbol{\alpha}\left(\exp\{\mathbf{D}t\} - \mathbf{I}_m\right)\left(\mathbf{D} - \mathbf{e}_m\boldsymbol{\theta}\right)^{-1}\overline{\mathbf{D}}_1\mathbf{e}_m, \quad t \geq 0.$$

Since $\boldsymbol{\alpha}\exp\{\mathbf{D}t\}$ converges to $\boldsymbol{\theta}$ as $t \to \infty$ (see Latouche and Ramaswami (1999)), we find that $\lim_{t\to\infty} E[N(t)]/t = \lambda$, so λ is the expected number of arrivals per unit time.

If the initial phase vector is $\boldsymbol{\theta}$ (i.e., we set $\boldsymbol{\alpha} = \boldsymbol{\theta}$), the Palm function reduces to $E[N(t)] = \lambda t$. For the variance of the number of arrivals in $(0,t]$ and the covariance of the counts, we refer to the results summarized in Subsection 2.3; see also Narayana and Neuts (1992).

2.2.2. Superposition and thinning

The class of BMAPs is closed under superposition. For simplicity, we consider two independent BMAPs $\{(N_i(t), J_i(t)); t \geq 0\}$ with characteristic sequences $\{\mathbf{D}_k^i; k \geq 0\}$ of order m_i, for $i \in \{1,2\}$, but the construction can be readily extended to an arbitrary number of BMAPs. Then, the resulting superposition process $\{(N(t), J(t)); t \geq 0\}$ is a BMAP with matrices $\{\mathbf{D}_k^1 \oplus \mathbf{D}_k^2; k \geq 0\}$. We notice that the count $N(t)$ is defined by $N_1(t) + N_2(t)$ and the phase process $J(t)$ has the form $(J_1(t), J_2(t))$.

Thinning is a mechanism to split or remove a part of the arrivals generated by the BMAP. As a result, thinning can be thought of as an operation opposite to the superposition. One way to single out arrivals from the original BMAP flow is just to discard any individual arrival with probability p independently of the rest of arrivals. The resulting BMAP has a matrix representation $\{\mathbf{D}_k^T; k \geq 0\}$, where

$$\mathbf{D}_0^T = \mathbf{D}_0 + \sum_{j=1}^{\infty} p^j \mathbf{D}_j,$$

$$\mathbf{D}_k^T = \sum_{j=k}^{\infty} \binom{j}{k} p^{j-k}(1-p)^k \mathbf{D}_j, \quad k \geq 1.$$

Another more sophisticated way to understand the thinning is associated with the arrivals of a BMAP and a clock with a PH distribution with representation $(\boldsymbol{\tau}, \mathbf{T})$. An auxiliary state 0 indicates that the PH clock is active, so that during this period the BMAP arrivals are not registered. As soon as the clock expires, the process turns to the auxiliary state 1 and the next arrival is registered. Immediately after one arrival is

registered, the PH clock is restarted. This description leads to a BMAP with matrices

$$\mathbf{D}_0^T = \begin{pmatrix} \mathbf{D} \oplus \mathbf{T} & \mathbf{I}_m \otimes \mathbf{t} \\ \mathbf{0}_{m \times mn} & \mathbf{D}_0 \end{pmatrix} \qquad \mathbf{D}_k^T = \begin{pmatrix} \mathbf{0}_{mn} & \mathbf{0}_{mn \times m} \\ \mathbf{D}_k \otimes \boldsymbol{\tau} & \mathbf{0}_{m \times m} \end{pmatrix}, \quad k \geq 1.$$

Decomposition of BMAPs provides another related operation. We may decompose a BMAP into n types of arrivals by considering independent markings with probabilities p_i, for $1 \leq i \leq n$, where $\sum_{i=1}^{n} p_i = 1$. Then, the split process $\{(N_i(t), J(t)); t \geq 0\}$ is a BMAP with $\mathbf{D}_0^i = \mathbf{D}_0 + (1-p_i)\overline{\mathbf{D}}_0$, $\mathbf{D}_k^i = p_i \mathbf{D}_k$, for $k \geq 1$ and each $1 \leq i \leq n$, where $\overline{\mathbf{D}}_0 = \mathbf{D} - \mathbf{D}_0$.

2.2.3. Local poissonification of a MAP

The local poissonification (see Neuts *et al.* (1992)) is an approach to quantifying the burstiness of a stationary point process. The events in successive intervals of length a are independently and uniformly redistributed over those intervals. The resulting local poissonification process mimics the behaviour of a PP over each interval.

For the MAP, the local poissonification construction can be tractably investigated by using matrix-analytic methods. To construct the stationary local poissonification of the MAP, we first choose the phase according to the vector $\boldsymbol{\theta}$ and a grid of points, regularly placed at a distance a. Then, the time origin is chosen randomly in one of the resulting intervals. Denote by $N_a(t)$ the counting process of the poissonification in any interval of length t.

The Palm function of $N_a(t)$ is $E[N_a(t)] = \lambda t$, for $t \geq 0$, thus showing that the poissonification preserves the fundamental rate of the original MAP. On the other hand, the variance of the count $N_a(t)$ is given by

$$\begin{aligned} Var(N_a(t)) = \lambda t + (V^0(a) - \lambda a) \left(\left(\frac{t}{a}\right)^2 - \frac{1}{3}\left(\frac{t}{a}\right)^3 + \frac{1}{3}\left(\frac{t-a}{a}\right)^3 V(t-a) \right) \\ + \frac{1}{3}\sum_{k=0}^{\infty} \rho_{k+1}(a) \left(\left(\frac{t-ka}{a}\right)^3 V(t-ka) - 2\left(\frac{t-(k+1)a}{a}\right)^3 V(t-(k+1)a) \right. \\ \left. + \left(\frac{t-(k+2)a}{a}\right)^3 V(t-(k+2)a) \right), \end{aligned}$$

where $V(x) = 1$ if $x \geq 0$, and it equals 0 otherwise, whereas $V^0(a)$ and $\rho_k(a)$ denote respectively the variance of the number of events in $(0,a]$ and the covariance of the counts in the intervals $(0,a]$ and $(ka,(k+1)a]$, in the stationary given MAP; see Subsection 2.3.1.

A number of computationally implementable descriptors include the dispersion function and the exponential peakedness (see Subsections 2.3.1 and 2.3.3), as well as

the distribution of the interval length. The latter is defined as the probability distribution of the interval between an arbitrary point and the next event in the poissonification of the stationary MAP. Its Laplace-Stieltjes transform $\varphi_a(s)$ is given by

$$\varphi_a(s) = 1 - \frac{s}{\lambda} + \frac{s^2 a}{\lambda}\left(\boldsymbol{\theta}\mathbf{L}_a^1(s)\mathbf{e}_m + \boldsymbol{\theta}L_a^0(s)\left(\mathbf{I}_m - e^{-sa}\exp\{\mathbf{D}_0 a\}\right)^{-1}\mathbf{L}_a^0(s)\mathbf{e}_m\right),$$

where the matrices $\mathbf{L}_a^0(s)$ and $\mathbf{L}_a^1(s)$ are defined by

$$\mathbf{L}_a^0(s) = \int_0^1 \exp\{\mathbf{D}^*(u)a\}e^{-sa(1-u)}du,$$

$$\mathbf{L}_a^1(s) = \int_0^1 u\exp\{\mathbf{D}^*(u)a\}e^{-sa(1-u)}du.$$

The mean μ_a and the variance σ_a^2 of the inter-arrival time are given by

$$\mu_a = \frac{1}{\lambda},$$

$$\sigma_a^2 = \frac{2a}{\lambda}\left(\boldsymbol{\theta}\mathbf{L}_a^1(0)\mathbf{e}_m + \boldsymbol{\theta}\mathbf{L}_a^0(0)\left(\mathbf{I}_m - \exp\{\mathbf{D}_0 a\}\right)^{-1}\mathbf{L}_a^0(0)\mathbf{e}_m\right) - \frac{1}{\lambda^2}.$$

2.2.4. Denseness property

Asmussen and Koole (1993) prove that a general class of *marked point processes* (MPP) can be approximated by appropriate MAPs. The MPP can be considered either at an arbitrary time or at selected discrete epochs. In the latter case the MPP is represented as a bivariate process $\{(T_n, Y_n); n \geq 0\}$, where the random variables T_n denote inter-arrival times and the marks Y_n are allowed to vary in $(0,\infty)$. In the arbitrary time version, an MPP is viewed as a point process taking values on the state space $[0,\infty)\times(0,\infty)$. A class of *Markovian arrival streams* (MAS) is also defined to approximate the given MPP. In a MAS there exists a finite state space of phases modulated by two matrices playing the same role that $\mathbf{D}_0$ and $\mathbf{D}_1$ in the MAP. When an arrival occurs, a mark is assigned according to a distribution B_{ij} on $(0,\infty)$. The mark depends on the current phase i and the destination phase j. If all B_{ij} are degenerate at 1, then the MAS agrees with the MAP.

The main result in Asmussen and Koole (1993) establishes that the class of MASs is dense in the class of MPPs in both time scales. The convergence must be viewed in distribution. However, related results for stationary processes and convergence of the moments also hold. It is interesting to remark that the convergence result does not hold when the class of MASs is replaced by MMPPs.

The above property is the analogue of the denseness property of PH distributions in the set of all probability distributions on $[0,\infty)$; see Neuts (1989). The proof follows from the fact that any probability distribution on $[0,\infty)$ may be suitably approximated by

a discrete distribution with a finite support, which is indeed a discrete PH distribution; see Latouche and Ramaswami (1999, Section 2.5) and Neuts (1981, Section 2.2).

2.3. Some interesting descriptors

The quantification of the main quality characteristics of the BMAP is of primarily theoretical and practical utility. This important objective is reached through the consideration of a variety of computationally implementable descriptors.

We distinguish three categories of descriptors for BMAPs: (a) descriptors associated with the counting function, (b) descriptors associated with inter-arrival times, and (c) other descriptors.

2.3.1. Descriptors associated with the counting function

To begin with, we recall that expressions for the fundamental arrival rate λ and the expected number of arrivals $E[N(t)]$ were already given in preceding subsections. Other descriptors related to the counting function are

(i) *The variance of the number of arrivals*. Given the initial distribution $\boldsymbol{\theta}$, we have

$$\begin{aligned} Var(N(t)) = &\left(\lambda_2 - 2\lambda^2 - 2\boldsymbol{\theta}\overline{\mathbf{D}}_1(\mathbf{D}-\mathbf{e}_m\boldsymbol{\theta})^{-1}\overline{\mathbf{D}}_1\mathbf{e}_m\right)t \\ &+ 2\boldsymbol{\theta}\overline{\mathbf{D}}_1(\mathbf{D}-\mathbf{e}_m\boldsymbol{\theta})^{-1}\left(\exp\{\mathbf{D}t\}-\mathbf{I}_m\right)(\mathbf{D}-\mathbf{e}_m\boldsymbol{\theta})^{-1}\overline{\mathbf{D}}_1\mathbf{e}_m, \end{aligned}$$

where $\lambda_2 = \boldsymbol{\theta}\overline{\mathbf{D}}_2\mathbf{e}_m$ and $\overline{\mathbf{D}}_2 = \sum_{k=1}^{\infty} k^2\mathbf{D}_k$.

(ii) *The dispersion function*. It is defined as

$$F_d(t) = \frac{Var(N(t))}{E[N(t)]}.$$

We observe that the dispersion function is a minor variant of the coefficient of variation, which is defined as the ratio between the standard deviation and the expectation. The dispersion function is also known as the index of dispersions for the counts; see Chakravarthy (2001).

(iii) *The covariance and the correlation of the counts*. Given the positive real numbers t, u, r and s, we construct the time intervals $(t,t+u]$ and $(t+u+r,t+u+r+s]$. The stationary versions of the covariance $\varphi(u,s,r)$ and the correlation $\rho(u,s,r)$ in these intervals are given by

$$\begin{aligned} \varphi(u,s,r) &= \boldsymbol{\theta}\overline{\mathbf{D}}_1(\mathbf{D}-\mathbf{e}_m\boldsymbol{\theta})^{-1}(\exp\{\mathbf{D}u\}-\mathbf{I}_m)\exp\{\mathbf{D}r\}(\exp\{\mathbf{D}s\}-\mathbf{I}_m) \\ &\quad\times(\mathbf{D}-\mathbf{e}_m\boldsymbol{\theta})^{-1}\overline{\mathbf{D}}_1\mathbf{e}_m - \lambda^2 us, \\ \rho(u,s,r) &= \frac{\varphi(u,s,r)}{\sqrt{Var(N(u))Var(N(s))}}. \end{aligned}$$

Those readers interested in the derivation of the above formulas are referred to the papers by Narayana and Neuts (1992), and Neuts *et al*. (1992).

2.3.2. Descriptors associated with inter-arrival times

Assume that $J(0)$ has a distribution $\boldsymbol{\alpha}$. The random vector $(\tau_1,\ldots,\tau_n)$ of inter-arrival times follows a multivariate continuous PH distribution (see Kulkarni (1989)). Therefore, the *n*th inter-arrival time τ_n has a PH distribution with representation

$$\left(\boldsymbol{\alpha}\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1},\mathbf{D}_0\right).$$

Then, it is immediate to obtain the expressions for the mean and the variance in the list below.

(i) *The mean of* τ_n

$$E[\tau_n]=\boldsymbol{\alpha}\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1}\left(-\mathbf{D}_0^{-1}\right)\mathbf{e}_m,\quad n\geq 1.$$

(ii) *The variance of* τ_n

$$Var(\tau_n)=2\boldsymbol{\alpha}\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1}\left(-\mathbf{D}_0\right)^{-2}\mathbf{e}_m-\left(\boldsymbol{\alpha}\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1}\left(-\mathbf{D}_0^{-1}\right)\mathbf{e}_m\right)^2,\quad n\geq 1.$$

(iii) *The coefficient of variation*

$$cv(\tau_n)=\frac{\sqrt{Var(\tau_n)}}{E[\tau_n]},\quad n\geq 1.$$

(iv) *The covariance and the correlation between* τ_1 *and* τ_n

$$\begin{aligned}\varphi(\tau_1,\tau_n)&=\boldsymbol{\alpha}\left(-\mathbf{D}_0^{-1}\right)\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1}\left(-\mathbf{D}_0^{-1}\right)\mathbf{e}_m\\&\quad-\left(\boldsymbol{\alpha}(-\mathbf{D}_0^{-1})\mathbf{e}_m\right)\left(\boldsymbol{\alpha}\left((-\mathbf{D}_0^{-1})\overline{\mathbf{D}}_0\right)^{n-1}(-\mathbf{D}_0^{-1})\mathbf{e}_m\right),\quad n\geq 1,\\\rho(\tau_1,\tau_n)&=\frac{\varphi(\tau_1,\tau_n)}{\sqrt{Var(\tau_1)Var(\tau_n)}}.\end{aligned}$$

Setting $\boldsymbol{\alpha}=\hat{\lambda}^{-1}\boldsymbol{\theta}\overline{\mathbf{D}}_0$, we obtain simplified expressions for the mean $\mu=\hat{\lambda}^{-1}$, the variance $\sigma^2=2\mu\boldsymbol{\theta}\left(-\mathbf{D}_0^{-1}\right)\mathbf{e}_m-\mu^2$ and the correlation

$$\rho(\tau_1,\tau_n)=\frac{\mu\boldsymbol{\theta}\left(\left(-\mathbf{D}_0^{-1}\right)\overline{\mathbf{D}}_0\right)^{n-1}\left(-\mathbf{D}_0^{-1}\right)\mathbf{e}_m-\mu^2}{\sigma^2},$$

where $\hat{\lambda}$ is the batch arrival rate defined by $\hat{\lambda} = \boldsymbol{\theta}\overline{\mathbf{D}}_0\mathbf{e}_m$; see Neuts (1995). Thus, $\boldsymbol{\alpha} = \hat{\lambda}^{-1}\boldsymbol{\theta}\overline{\mathbf{D}}_0$ represents the stationary distribution of the phase right after the arrival of a batch.

Example 2.2 We illustrate here the computation of the inter-arrival descriptors for the BMAP described in Example 2.1. The stationary probability vector $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta} = (8/17, 5/17, 2/17, 2/17)$. Then, the arrival rates $\lambda_k = \boldsymbol{\theta}\mathbf{D}_k\mathbf{e}_4$ of batches of size k, for $k \in \{1,2\}$, are given by $\lambda_1 = 1.0$ and $\lambda_2 = 1.17647$, while the batch and the total arrival rates are given by $\hat{\lambda} = \lambda_1 + \lambda_2$ and $\lambda = \lambda_1 + 2\lambda_2$, respectively.

By taking $\boldsymbol{\alpha} = \hat{\lambda}^{-1}\boldsymbol{\theta}\overline{\mathbf{D}}_0$, we easily obtain the values $E[\tau_1] = 0.45945$, $Var(\tau_1) = 0.48619$, $\varphi(\tau_1, \tau_5) = 0.00832$ and $\rho(\tau_1, \tau_5) = 0.01711$.

2.3.3. Other descriptors

(i) *Peakedness*. The peakedness functional is a second order descriptor used in communication engineering. It is a functional of the holding time distribution defined as the ratio between the variance and the expectation of the number of busy servers in a queue with infinite servers and independent, identically distributed service times, which is feeded by a certain arrival process. The particular case where the service times are exponentially distributed with rate $\mu > 0$ is called the exponential peakedness.
Eckberg (1983) has shown that the exponential peakedness $z_{exp}(\mu)$ and the Laplace-Stieltjes transform $\phi_{arr}(s)$ of the expected number of arrivals in $(0,t]$, starting from an arbitrary arrival, are related by the formula

$$z_{exp}(\mu) = 1 + \phi_{arr}(\mu) - \frac{\lambda}{\mu}.$$

Following Neuts *et al.* (1992), we observe that the exponential peakedness for the MAP is obtained from the explicit formulas for the kth factorial moments of the number of customers in the $MAP/M/\infty$ queue, which are given by

$$\mathbf{f}_k = k!\boldsymbol{\theta}\mathbf{D}_1(\mu\mathbf{I}_m - \mathbf{D})^{-1}\mathbf{D}_1(2\mu\mathbf{I}_m - \mathbf{D})^{-1}\cdots\mathbf{D}_1(k\mu\mathbf{I}_m - \mathbf{D})^{-1}, \quad k \geq 1.$$

Thus, we have

$$z_{exp}(\mu) = \frac{\mathbf{f}_2\mathbf{e}_m + \mathbf{f}_1\mathbf{e}_m - (\mathbf{f}_1\mathbf{e}_m)^2}{\mathbf{f}_1\mathbf{e}_m}.$$

For the exponential peakedness of the local poissonification of the MAP, we refer the reader to Neuts *et al.* (1992).

(ii) *Index of burstiness*. The term burstiness is referred to an arrival process whose flow exhibits short intervals with a large number of arrivals separated by long

intervals with few arrivals. In order to quantify burstiness, Neuts (1993) proposed to thinning the original arrival process with the help of an auxiliary labeling process.

Assume that the arrival process is a *Markov renewal process* (MRP) whose Markov renewal sequence has a kernel $\mathbf{H}(x) = (h_{ij}(x))$, where the transition probabilities $p_{ij} = h_{ij}(\infty)$ take values on the finite set $\{1, \ldots, r\}$; see Kulkarni (1995). We choose the labeling process to be a stationary MAP independent of the MRP. A point of the MRP is registered if and only if it is immediately preceded by an arrival of the labeling MAP. If the fundamental rate λ decreases, typically only a few arrivals of the MRP are registered. More importantly, the MRP arrivals occurring in intense short runs are most likely to be unregistered. Thus, the proposed labeling mechanism removes the bursts of the MRP.

Suppose that, in the stationary version of the MRP, arrivals occur at rate δ. Let $\boldsymbol{\pi}$ be the invariant distribution of the stochastic matrix $\mathbf{H}(\infty) = (p_{ij})$. Then, we define the index $\chi(p)$ of burstiness by

$$\chi(p) = \frac{1}{\delta}\kappa^{-1}(p), \quad 0 \le p \le 1,$$

where $\kappa^{-1}(p)$ is the inverse function of $\kappa(\lambda)$ defined by

$$\kappa(\lambda) = 1 - \int_0^\infty \boldsymbol{\theta} \exp\{\mathbf{D}_0 u\}\mathbf{e}_m d\,(\boldsymbol{\pi}\mathbf{H}(u)\mathbf{e}_r)\,.$$

Thus, $\delta\chi(p)$ is interpreted as the rate of the MAP labeling process for which a fraction p of the arrivals of the MRP are registered.

In Neuts (1993), the analysis is even extended to investigate correlations and run distributions.

We conclude this subsection by illustrating the calculation of $\chi(p)$ for the *interrupted Poisson process* (IPP).

Example 2.3 An IPP is a bursty MAP with $m = 2$ and matrices

$$\mathbf{D}_0 = \begin{pmatrix} -(\lambda_a + \delta_1) & \delta_1 \\ \delta_2 & -\delta_2 \end{pmatrix}, \qquad \mathbf{D}_1 = \begin{pmatrix} \lambda_a & 0 \\ 0 & 0 \end{pmatrix}.$$

This means that a PP of rate λ_a can be interrupted with probability $\delta_1(\lambda_a + \delta_1)^{-1}$. If this occurs, then an interruption period (exponentially distributed of rate δ_2) takes place.

Assume that the MAP labeling process is Poisson of rate λ. By using the fact that the IPP is equivalent to a certain hyperexponential renewal process (see Milne (1982)), it is easy to find that

$$\kappa(\lambda) = 1 - \frac{\lambda_a(\lambda+\delta_2)}{\lambda^2+\lambda(\lambda_a+\delta_1+\delta_2)+\lambda_a\delta_2}.$$

By normalizing the fundamental rate of the IPP to be one, we obtain the following expression for the index of burstiness:

$$\chi(p) = \frac{p\rho^{-1}-\bar{p}\sigma+\sqrt{(p\rho^{-1}-\bar{p}\sigma)^2+4p\bar{p}\sigma}}{2\bar{p}}, \quad 0<p<1,$$

where $\bar{p}=1-p$, $\sigma=\delta_1+\delta_2$ and $\rho=\delta_2/\sigma$.

2.4. Some applications

The next examples in queueing, reliability and inventory models are intended to help the reader acquire some feeling for the range of applications of the BMAP and its variants. By means of them, we briefly motivate the use of structured Markov chains; see Bini *et al.* (2005), Latouche and Ramaswami (1999), Li (2010) and Neuts (1981,1989).

2.4.1. The BMAP/G/1 queue

Consider a single-server queue whose arrival process is a BMAP with sequence $\{\mathbf{D}_k; k\geq 0\}$. Let the service times have an arbitrary probability distribution function $H(x)$.

We may find many similarities between the $BMAP/G/1$ and the $M/G/1$ queues. To begin with, we construct an embedded Markov chain $\{(Q_n,J_n); n\geq 0\}$ at the times of service completions by defining the pair (Q_n,J_n) as the queue length and the phase of the BMAP immediately after the nth service completion. Define the matrices

$$\mathbf{A}_n = \int_0^\infty \mathbf{P}(n,u)dH(u), \quad n\geq 0,$$

$$\mathbf{B}_n = \sum_{k=1}^{n+1}\int_0^\infty \exp\{\mathbf{D}_0 u\}du\mathbf{D}_k\int_0^\infty \mathbf{P}(n+1-k,v)dH(v)$$

$$= -\mathbf{D}_0^{-1}\sum_{k=1}^{n+1}\mathbf{D}_k\mathbf{A}_{n+1-k}, \quad n\geq 0.$$

The matrix $\mathbf{A}_n=(a_{ij}(n))$ consists of the conditional probabilities that n customers arrive during a service time starting from phase i and finishing at phase j of the BMAP. We can therefore describe some of the transition probabilities for the embedded Markov chain by

$$P(Q_1=l+n-1, J_1=j \,|\, Q_0=l, J_0=i) = a_{ij}(n), \quad n\geq 0, 1\leq i,j\leq m,$$

independently of $l \geq 1$. It can be readily verified that the matrix generating function $\mathbf{A}^*(z) = \sum_{n=0}^{\infty} z^n \mathbf{A}_n$ is given by

$$\mathbf{A}^*(z) = \int_0^{\infty} \exp\{\mathbf{D}^*(z)u\} dH(u).$$

Similarly, the matrix $\mathbf{B}_n = (b_{ij}(n))$ contains the probabilities that first a batch of k customers arrives and then $n+1-k$ additional customers arrive during the subsequent service time, for $1 \leq k \leq n+1$. Note that this situation occurs whenever a service completion leaves the queue empty. Hence, we can write down

$$P(Q_1 = n, J_1 = j | Q_0 = 0, J_0 = i) = b_{ij}(n), \quad n \geq 0, 1 \leq i, j \leq m.$$

As a result, the one-step transition probability matrix of $\{(Q_n, J_n); n \geq 0\}$ is given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \dots \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \dots \\ & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \dots \\ & & \mathbf{A}_0 & \mathbf{A}_1 & \dots \\ & & & \ddots & \ddots \end{pmatrix}.$$

A matrix of this structured form is said to be of $M/G/1$-type (see Neuts (1989)), which underlines the similarity to the univariate embedded Markov chain of the $M/G/1$ queue.

The $BMAP/G/1$ was first analyzed in Ramaswami (1980), where the BMAP was used under its older, more complicated notation. An outline of Ramaswami's results under the present matrix formulation, along with some new results, are presented in Lucantoni (1991). For a historical survey on the model, see Lucantoni (1993).

2.4.2. The D-BMAP/D/1/K queue

Consider a discrete-time queue in which arrivals are generated by M independent input sources. Incoming arrivals are queued in a shared buffer of capacity K, with $K < M$. The time needed to serve an arrival is selected as time unit and named slot. Each input source in a slot takes either ON state or OFF state. When an input source is in ON state, one arrival is generated with probability g. If the source is in OFF state, then no arrival is generated. Suppose also that any OFF (or ON) source in a time slot changes to the ON (or OFF) state with probability p (or q) in the next slot. This superposition of sources can be modelled as a *discrete-time batch Markovian arrival process* (D-BMAP); see Subsection 3.1.

Let Q_n and J_n be the queue length and the number of ON sources (phase) at the nth slot. Then, the sequence $\{(Q_n, J_n); n \geq 1\}$ is a discrete-time Markov chain on the state space $\{0, 1, \dots, K\} \times \{0, 1, \dots, M\}$ with one-step transition probability matrix

$$\mathbf{P} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_{K-1} & \sum_{k=K}^{M} \mathbf{D}_k \\ \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_{K-1} & \sum_{k=K}^{M} \mathbf{D}_k \\ & \mathbf{D}_0 & \mathbf{D}_1 & \cdots & \mathbf{D}_{K-2} & \sum_{k=K-1}^{M} \mathbf{D}_k \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & & \mathbf{D}_0 & \sum_{k=1}^{M} \mathbf{D}_k \end{pmatrix},$$

where the matrices $\mathbf{D}_k$ have the following elements:

$$d_{ii'}(k) = \binom{i}{k} g^k (1-g)^{i-k} f_{ii'}, \quad 0 \le k \le i,$$

and $f_{ii'}$, for $0 \le i, i' \le M$, is given by

$$f_{ii'} = \sum_{j=0}^{i} \binom{i}{j} q^j (1-q)^{i-j} \binom{M-i}{i'+j-i} p^{i'+j-i} (1-p)^{M-i'-j}.$$

The binomial term in $d_{ii'}(k)$ is the probability of k arrivals in the current slot, given that the number of ON sources is i. On the other hand, $f_{ii'}$ is the probability that in the next slot there will be i' ON sources, given that in the current slot there are i.

The structure of $\mathbf{P}$ shows that $\{(Q_n, J_n); n \ge 1\}$ is a finite Markov chain of $M/G/1$-type. This structured Markov chain, but involving a more sophisticated sequence $\{\mathbf{D}_k; k \ge 0\}$, is the analytical model used by Blondia and Casals (1992) for a statistical multiplexer whose input consists of the superposition of *variable bit rate* (VBR) sources.

2.4.3. A reliability system subject to failures

Consider a system subject to internal and external failures. An internal failure causes a fatal failure of the system and implies that the system must be replaced. External failures affect the system in two ways: some of them cause damage that can be repaired, whereas others cause fatal failure and consequently the system must be replaced. Assume that the replacement and repair operations are instantaneous.

In practice, it is frequent that a system can bear only a certain number of failures, in such a way that when the next failure occurs it is replaced. Let $k \ge 1$ be the maximum number of imperfect repairs that the system can undergo. At an arbitrary time, the state of the system can be described by means of the number $K(t)$ of imperfect repairs suffered by the system in process at time t. The random variable $K(t)$ takes values in the set $\{0, 1, \ldots, k\}$ and, in particular, it records the state 0 if the system in process at time t is new.

Montoro-Cazorla and Pérez-Ocón (2006) use a matrix-analytic approach when the lifetime of the system due to wear out follows a PH distribution, with representation

$(\boldsymbol{\tau}, \mathbf{T})$ of order n. Arrivals of external failures are modelled by a MMAP (see Subsection 3.2) with two types of marks referring to external failures with minimal repair and external failures causing a replacement. In the characteristic matrices $\{\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2\}$ of dimension m, the matrix $\mathbf{D}_1$ refers to the occurrence of an external failure with minimal repair, and $\mathbf{D}_2$ refers to a failure that causes the replacement of the system. The matrix $\mathbf{D}_0$ records those changes that do not imply any failure.

Then, a Markovian description of the system state follows from the Markov chain $\{(K(t), J_l(t), J_a(t)); t \geq 0\}$, where $J_l(t)$ and $J_a(t)$ denote the lifetime phase and the phase of the arrival process, respectively, at time t. This is a Markov chain on the space state $\{0, 1, \ldots, k\} \times \{1, \ldots, n\} \times \{1, \ldots, m\}$ and infinitesimal generator

$$\mathbf{Q} = \begin{pmatrix} (\mathbf{T} + \mathbf{t}\boldsymbol{\tau}) \oplus \mathbf{D}_0 + \mathbf{e}_n\boldsymbol{\tau} \otimes \mathbf{D}_2 & \mathbf{I}_n \otimes \mathbf{D}_1 & & & \\ \mathbf{t}\boldsymbol{\tau} \otimes \mathbf{I}_m + \mathbf{e}_n\boldsymbol{\tau} \otimes \mathbf{D}_2 & \mathbf{T} \oplus \mathbf{D}_0 & \mathbf{I}_n \otimes \mathbf{D}_1 & & \\ \vdots & & \ddots & \ddots & \\ \mathbf{t}\boldsymbol{\tau} \otimes \mathbf{I}_m + \mathbf{e}_n\boldsymbol{\tau} \otimes \mathbf{D}_2 & & & \mathbf{T} \oplus \mathbf{D}_0 & \mathbf{I}_n \otimes \mathbf{D}_1 \\ \mathbf{t}\boldsymbol{\tau} \otimes \mathbf{I}_m + \mathbf{e}_n\boldsymbol{\tau} \otimes (\mathbf{D}_1 + \mathbf{D}_2) & & & & \mathbf{T} \oplus \mathbf{D}_0 \end{pmatrix}.$$

Therefore, the structural form of $\mathbf{Q}$ yields a finite Markov chain of $GI/M/1$-type; see Neuts (1981).

2.4.4. A multi-location inventory system

The next example (see Ching (1997)) is an inventory system in a multi-location situation under continuous review and one-for-one replenishment.

Consider a multi-location inventory system consisting of K locations that replenish their stocks from a common main depot. For the ith location, the inventory system is modelled by the $M/M/s_i/q_i$ queue with arrival rate λ_i and exponentially distributed lead times of each server with parameter μ_i. The overflow process of demand of the ith location can be approximated by a two-state MMPP with underlying matrices

$$\mathbf{Q}_{ia} = \begin{pmatrix} -\sigma_{i1} & \sigma_{i1} \\ \sigma_{i2} & -\sigma_{i2} \end{pmatrix}, \qquad \boldsymbol{\Lambda}_i = \begin{pmatrix} \lambda_i & 0 \\ 0 & 0 \end{pmatrix}.$$

The first state is equivalent to the event $\{$*the ith location is full*$\}$, and the second one amounts to the event $\{$*the ith location is not yet full*$\}$. Note that, in the former case, the maximum level of backlogs is attained and, consequently, a further demand will overflow to the main depot whenever the queue remains full. In the latter case, a further demand will be acceptable. Based on the stationary distribution of the $M/M/s_i/q_i$ queue, the parameters σ_{i1} and σ_{i2} are approximated as $\sigma_{i1} = s_i\mu_i$ and $\sigma_{i2} = b_i s_i \mu_i/(1 - b_i)$, where b_i denotes the blocking probability at the ith location

$$b_i = \sum_{j=-q_i}^{s_i} \prod_{k=1}^{s_i-j} \frac{\lambda_i}{\mu_i \min(k, s_i)}.$$

Therefore, we may regard the $MMPP/M/s/q$ queue describing the inventory system at the main depot as a finite Markov chain $\{(Q(t), J(t)); t \geq 0\}$ on the state space $\{-q, \ldots, s\} \times \{1, \ldots, 2^K\}$, where $Q(t)$ is the inventory level at the depot and $J(t)$ is the phase of the underlying Markov chain with infinitesimal generator $\mathbf{Q}_a = \mathbf{Q}_{1a} \oplus \ldots \oplus \mathbf{Q}_{Ka}$. Negative values for the inventory level $Q(t)$ amount to backlog.

The infinitesimal generator $\mathbf{Q}$ of $\{(Q(t), J(t)); t \geq 0\}$ has the following structured form:

$$\begin{pmatrix}
\mathbf{Q}_a - \mathbf{\Lambda} & \mathbf{\Lambda} & & & & & & \\
\mu \mathbf{I}_{2^K} & \mathbf{Q}_a - \mathbf{\Lambda} - \mu \mathbf{I}_{2^K} & \mathbf{\Lambda} & & & & & \\
 & \ddots & \ddots & \ddots & & & & \\
 & & s\mu \mathbf{I}_{2^K} & \mathbf{Q}_a - \mathbf{\Lambda} - s\mu \mathbf{I}_{2^K} & \mathbf{\Lambda} & & & \\
 & & & s\mu \mathbf{I}_{2^K} & \mathbf{Q}_a - \mathbf{\Lambda} - s\mu \mathbf{I}_{2^K} & \mathbf{\Lambda} & & \\
 & & & & \ddots & \ddots & & \\
 & & & & & s\mu \mathbf{I}_{2^K} & \mathbf{Q}_a - s\mu \mathbf{I}_{2^K}
\end{pmatrix},$$

where $\mathbf{\Lambda} = \mathbf{\Lambda}_1 \oplus \ldots \oplus \mathbf{\Lambda}_K$.

The stationary distribution of $\mathbf{Q}$ can be readily derived from the general theory of finite QBD processes; see e.g. Latouche and Ramaswami (1999, Chapter 10). For more information on finite QBD processes arising in manufacturing problems, the reader is referred to the monograph by Ching (2001).

3. Variants and extensions of the BMAP

In this section we collect several generalizations and variants of the BMAP. We start in Subsection 3.1 by presenting the D-BMAP; that is, the discrete-time analogue of the BMAP. The use of discrete-time models is motivated by many applications in communication systems where the basic units are digital. The consideration of Markov arrival processes with marked transitions opens new directions to investigate stochastic models with multiple types of items, fluid input, spatial arrivals, etc. In Subsection 3.2 we follow the original formulation by He and Neuts (1998) to introduce the MMAP. The HetSigma approach summarized in Subsection 3.3 provides a versatile way to get joint modulation of the arrival and service processes. In Subsection 3.4, under the title Markov-additive arrival processes, we briefly introduce some generalized arrival processes which allow the counting/marked and background processes to take values on more general spaces. The time-inhomogeneous case and the possibility of incorporating spatial features can also be subsumed under appropriate versions of the

Markov-additive umbrella. Finally, in Subsection 3.5 we deal with the BSDE approach which has been recently presented by Artalejo and Gómez-Corral (2010) as a tool for constructing Markov modulated stochastic models taking into account the reduction of dimensionality inherent to the matrix formulation.

3.1. The D-BMAP

The D-BMAP was introduced by Blondia and Casals (1992) as the discrete-time analogue of the BMAP. They showed that many useful discrete-time arrival processes can be obtained as particular cases of the D-BMAP and how this versatile arrival pattern can be used as *asynchronous transfer mode* (ATM) source model.

The key point in the constructive description of the D-BMAP is the consideration of finite matrices $\{\mathbf{D}_k; k \geq 0\}$, which govern phase transitions and batch sizes. Suppose that at time k the phase in progress is i, for $1 \leq i \leq m$. At the next time epoch $k+1$, a transition to another or the same phase takes place and a batch arrival may occur or not. More concretely, the elements $d_{ij}(0)$ of matrix $\mathbf{D}_0$ give the probabilities that the phase goes to state j with no arrival, given that the initial phase is i. On the other hand, the elements $d_{ij}(k)$ of $\mathbf{D}_k$ denote that, in the next time unit, there is a transition from phase i to phase j with a batch of size $k \geq 1$. We notice that

$$\sum_{j=1}^{m} \sum_{k=0}^{\infty} d_{ij}(k) = 1, \quad 1 \leq i \leq m.$$

We also assume that the matrix $\mathbf{I}_m - \mathbf{D}_0$ is non-singular, so the D-BMAP has an arrival with probability one.

With the help of $\{\mathbf{D}_k; k \geq 0\}$, we formally define the D-BMAP as the bivariate process $\{(N_k, J_k); k \geq 0\}$, where $\{J_k; k \geq 0\}$ is the background phase Markov chain and N_k denotes the counting variable. The one-step transition probability matrix of the D-BMAP is given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \mathbf{D}_3 & \cdots \\ & \mathbf{D}_0 & \mathbf{D}_1 & \mathbf{D}_2 & \cdots \\ & & \mathbf{D}_0 & \mathbf{D}_1 & \cdots \\ & & & \ddots & \ddots \end{pmatrix}.$$

A number of well-known processes are obtained by choosing appropriately the sequence of matrices $\{\mathbf{D}_k; k \geq 0\}$. The list includes the *Bernoulli arrival process*, the *Markov modulated Bernoulli process*, the *batch Bernoulli process with correlated arrivals* and many other processes which, in general, can be considered as the discrete counterparts of those particular cases of the BMAP listed in Subsection 2.1. For further details of other special cases of the D-BMAP, we refer to the papers by Chakravarthy (2001,2010).

We also remark that, like in the continuous-time BMAP, many interesting properties (such as counting, descriptors, superpositions, etc.) can be investigated. Since arguments are similar, these results will not be presented here, but we refer to the paper by Chakravarthy (2010) for a summary of basic results for the D-BMAP.

In what follows, we focus on the class of *platoon arrival processes* (PAP).

Example 3.1 The following description of the PAP is based on the paper by Alfa and Neuts (1995), who used the PAP to model vehicular traffic. Recently, Breuer and Alfa (2005) used a terminating D-MAP to generalize the concept of PAP.

The PAP is a discrete-time arrival process composed of platoons. Suppose that the number of arrivals in a platoon is a discrete PH of order d with representation $(\boldsymbol{\delta}, \mathbf{D})$ and absorption vector $\mathbf{d}$. Moreover, we assume that $p_1 = \delta_0 = 1 - \boldsymbol{\delta}\mathbf{e}_d > 0$ is the probability of a platoon consisting of a single vehicle (i.e., the probability of starting in the absorbing state) and $p_k = \boldsymbol{\delta}\mathbf{D}^{k-2}\mathbf{d}$, for $k \geq 2$, is the probability of having k arrivals in the platoon. In a first general approach, intraplatoon intervals separating two arrivals in the same platoon, have the probability mass function $\{p_1(k); k \geq 1\}$. On the other hand, the interplatoon interval separating the last arrival in a platoon and the first one of the immediately following platoon, have the probability mass function $\{p_2(k); k \geq 1\}$.

Let S_n be the nth arrival epoch and suppose that Y_n records the phase of the discrete PH distribution observed at time S_n+, whose representation is given by $(\boldsymbol{\delta}, \mathbf{D})$. Then, the PAP is the MRP associated with the Markov renewal sequence $\{(Y_n, S_n); n \geq 0\}$, whose kernel is described by the matrices

$$\mathbf{H}(j) = \begin{pmatrix} \delta_0 p_2(j) & \boldsymbol{\delta} p_2(j) \\ \mathbf{d} p_1(j) & \mathbf{D} p_1(j) \end{pmatrix}, \quad j \geq 1.$$

For practical purposes, the MRP formalism can be simplified by assuming that the intraplatoon intervals and the interplatoon intervals are distributed as discrete PH distributions with representations $(\boldsymbol{\alpha}_i, \mathbf{T}_i)$ with m_i phases and absorption vectors $\mathbf{t}_i$, for $i \in \{1,2\}$, respectively. The vectors $\boldsymbol{\alpha}_i$, for $i \in \{1,2\}$, are now assumed to be probability vectors. Thus, the PAP can be now seen as a D-MAP with matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ given by

$$\mathbf{D}_0 = \begin{pmatrix} \mathbf{T}_2 & \mathbf{0}_{m_2 \times dm_1} \\ \mathbf{0}_{dm_1 \times m_2} & \mathbf{I}_d \otimes \mathbf{T}_1 \end{pmatrix}, \qquad \mathbf{D}_1 = \begin{pmatrix} \delta_0 \mathbf{t}_2 \boldsymbol{\alpha}_2 & \boldsymbol{\delta} \otimes \mathbf{t}_2 \boldsymbol{\alpha}_1 \\ \mathbf{d} \otimes \mathbf{t}_1 \boldsymbol{\alpha}_2 & \mathbf{D} \otimes \mathbf{t}_1 \boldsymbol{\alpha}_1 \end{pmatrix},$$

where the underlying states (i, j) denote the phase of the discrete PH law with representation $(\boldsymbol{\delta}, \mathbf{D})$ and the phase of the (interplatoon or intraplatoon) interval in process.

3.2. The marked Markovian arrival process

The MMAP can be viewed as a multi-class extension of the BMAP. Although the analysis can be presented both in discrete- and continuous-time, we restrict our exposition

to the latter case. Similar to the BMAP, the MMAP definition is based on a background Markov chain $\{J(t); t \geq 0\}$, often called phase chain, with m states, which determines the arrivals of some marks taking values on a set $\mathscr{C}^0$. The set of marks $\mathscr{C}^0$ may have different interpretations, as we show in the sequel.

Let $\mathscr{C}^0$ be a finite or countable set of indices. More specifically, we may assume that a generic element $\mathbf{h}$ of $\mathscr{C}^0$ is a K-tuple $(h_1, \ldots, h_K)$, where $h_k \in \mathbb{N}$, for $1 \leq k \leq K$, and at least one coordinate is strictly positive. Define the non-negative matrices $\mathbf{D}_0$ and $\{\mathbf{D}_\mathbf{h}; \mathbf{h} \in \mathscr{C}^0\}$ of order m. The entries of $\mathbf{D}_0$ describe the motion of the phase Markov chain without any arrival. $\mathbf{D}_0$ is assumed to be a non-singular matrix with negative diagonal elements. The matrices $\mathbf{D}_\mathbf{h}$ are non-negative and give the transition rates of the phase Markov chain with a mark $\mathbf{h}$. Then, $\mathbf{D} = \mathbf{D}_0 + \sum_{\mathbf{h} \in \mathscr{C}^0} \mathbf{D}_\mathbf{h}$ is an infinitesimal generator. The counting process $\{(N_\mathbf{h}(t), J(t)); \mathbf{h} \in \mathscr{C}^0, t \geq 0\}$ is called a MMAP.

Alternatively, we may define the MMAP in terms of PPs. To this end, it is enough to replace the role of the rates $\{d_{ij}(k); 1 \leq i, j \leq m\}$, for $k \geq 1$, in Remark 2.2 by the analogue marked version $\{d_{ij}(\mathbf{h}); 1 \leq i, j \leq m\}$, for $\mathbf{h} \in \mathscr{C}^0$. The semi-Markovian representation in Remark 2.1 for the BMAP also holds for the MMAP.

It is clear that the choice $K = 1$ and $\mathscr{C}^0 = \mathbb{N} - \{0\}$ reduces the MMAP to the BMAP. The case $K = 1$ and $\mathscr{C}^0 = \{1, \ldots, C\}$ determines arrivals of C different types of customers or items; that is, the MMAP is interpreted as a proper multi-class generalization of the BMAP.

The following specifications of the matrices $\mathbf{D}_0$ and $\{\mathbf{D}_\mathbf{h}; \mathbf{h} \in \mathscr{C}^0\}$ show interesting features captured under the MMAP formulation:

(i) A reinterpretation of the batch sizes in terms of different classes of customers allows us to see example (ii) for cyclic arrivals in Section 2 as an arrival process where type-1 and type-2 customers arrive cyclically.

(ii) Individual vs group

$$\mathbf{D}_0 = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}, \quad \mathbf{D}_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D}_{2,1} = \begin{pmatrix} 0 & 0.5 \\ 1 & 0 \end{pmatrix}.$$

First, we notice that the marks $\mathscr{C}^0 = \{\{1\}, \{2,1\}\}$ can be put in correspondence with the case $K = 1$ and $\mathscr{C}^0 = \{1, 2\}$. This comment can be readily extended to any arbitrary finite set $\mathscr{C}^0$.

In this arrival process, there are individual arrivals of type-1 and group arrivals where the group consists of one type-2 customer accompanied by a type-1 customer.

(iii) Type-2 follows type-1

$$\mathbf{D}_0 = \begin{pmatrix} -4 & 0 \\ 0 & -5 \end{pmatrix}, \quad \mathbf{D}_1 = \begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{D}_{2,1} = \begin{pmatrix} 0 & 0 \\ 5 & 0 \end{pmatrix}.$$

A group arrival $\{2,1\}$ is always preceded by the arrival of a customer of type-1.

(iv) Orders within batches

$$\mathbf{D}_0 = \begin{pmatrix} -15 & 0 \\ 0 & -10 \end{pmatrix}, \quad \mathbf{D}_{\{112\}} = \begin{pmatrix} 14 & 0 \\ 0 & 9 \end{pmatrix}, \quad \mathbf{D}_{\{121\}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The marks $\{112\}$ and $\{121\}$ are associated with group arrivals of size 3. Each group consists of two type-1 customers and one customer of type-2. The orders in which individuals are scheduled within a group do matter, so the two marks are distinguished.

Among the descriptors of the MMAP, we stress the interest in the counting functions. The generating function of $\mathbf{N}(t) = (N_1(t), \ldots, N_K(t))$ is given by

$$\mathbf{P}^*(\mathbf{z},t) = \sum_{\mathbf{n}} \mathbf{z}^{\mathbf{n}} \mathbf{P}(\mathbf{n},t) \; = \; \exp\{\mathbf{D}^*(\mathbf{z})t\},$$

where $\mathbf{n} = (n_1, \ldots, n_K)$ with $n_i \geq 0$, for $1 \leq i \leq K$, and $\mathbf{P}(\mathbf{n},t)$ is the matrix with elements $P_{ij}(\mathbf{n},t) = P(\mathbf{N}(t) = \mathbf{n}, J(t) = j | \mathbf{N}(0) = \mathbf{0}_K, J(t) = i)$, while $\mathbf{z}^{\mathbf{n}} = z_1^{n_1} \cdots z_K^{n_K}$ and $\mathbf{D}^*(\mathbf{z}) = \mathbf{D}_0 + \sum_{\mathbf{h} \in \mathscr{C}^0} \mathbf{z}^{\mathbf{h}} \mathbf{D}_{\mathbf{h}}$, for $|z_k| \leq 1$ and $1 \leq k \leq K$.

Now the covariances and correlations between $\{N_{\mathbf{h}}(t); t \geq 0\}$, for $\mathbf{h} \in \mathscr{C}^0$, can be explicitly expressed; see He and Neuts (1998).

For easiness, we assume $\mathscr{C}^0 = \{1,2\}$; i.e., we have two types of arrivals.

Given any initial probability distribution $\boldsymbol{\alpha}$ for the phase Markov chain, we have

$$E[N_{\mathbf{h}}(t)] = \lambda_{\mathbf{h}} t + \boldsymbol{\alpha} (\exp\{\mathbf{D}t\} - \mathbf{I}_m) (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} \mathbf{D}_{\mathbf{h}} \mathbf{e}_m, \quad \mathbf{h} \in \mathscr{C}^0, \; t \geq 0,$$

where $\boldsymbol{\theta}$ is the stationary distribution of $\mathbf{D}$ and $\lambda_{\mathbf{h}} = \boldsymbol{\theta} \mathbf{D}_{\mathbf{h}} \mathbf{e}_m$ is the fundamental arrival rate of type-$\mathbf{h}$ marks.

If we take $\boldsymbol{\alpha} = \boldsymbol{\theta}$, then

$$\begin{aligned} Var(N_{\mathbf{h}}(t)) = & \left(\lambda_{\mathbf{h}} - 2\lambda_{\mathbf{h}}^2 - 2\boldsymbol{\theta} \mathbf{D}_{\mathbf{h}} (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} \mathbf{D}_{\mathbf{h}} \mathbf{e}_m \right) t \\ & + 2\boldsymbol{\theta} \mathbf{D}_{\mathbf{h}} (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} (\exp\{\mathbf{D}t\} - \mathbf{I}_m) (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} \mathbf{D}_{\mathbf{h}} \mathbf{e}_m, \end{aligned}$$

and the covariance between $N_1(t)$ and $N_2(t)$ is given by

$$\begin{aligned} \varphi(N_1(t), N_2(t)) = & - \left(2\lambda_1 \lambda_2 + \boldsymbol{\theta} \left(\sum_{k=1}^{2} \mathbf{D}_k (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} \mathbf{D}_{3-k} \right) \mathbf{e}_m \right) t \\ & + \boldsymbol{\theta} \left(\sum_{k=1}^{2} \mathbf{D}_k (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} (\exp\{\mathbf{D}t\} - \mathbf{I}_m) (\mathbf{D} - \mathbf{e}_m \boldsymbol{\theta})^{-1} \mathbf{D}_{3-k} \right) \mathbf{e}_m. \end{aligned}$$

We illustrate the computation of the counting moments by means of the BMAP considered in Examples 2.1 and 2.2. Obviously, the batch size becomes here the mark in the MMAP terminology.

Example 3.2 If the MMAP with matrices $\{\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2\}$ given in Example 2.1 is stationary, for $t = 2.5$, we get

$$E[N_1(t)] = 2.5, \qquad E[N_2(t)] = 2.94117,$$
$$Var(N_1(t)) = 4.24980, \qquad Var(N_2(t)) = 6.17905.$$

The covariance and correlation between $N_1(t)$ and $N_2(t)$ are given by

$$\varphi(N_1(t), N_2(t)) = 3.30791, \qquad \rho(N_1(t), N_2(t)) = 0.64551.$$

The mean and variance of the total number of counts $N(t) = N_1(t) + 2N_2(t)$ are $E[N(t)] = 8.38235$ and $Var(N(t)) = 42.19772$.

A good account of results for other basic properties of the MMAP, including thinning, type of arrivals, peakedness and closure properties, are found in He and Neuts (1998), and He (2010).

3.3. The HetSigma approach

The HetSigma approach (see Chakka and Do (2007)) has been proposed in order to evaluate the performance of queueing models with burstiness and correlation arising from applications to wireless broadband networks. The proposed modulation mechanism could be subsumed under a MMAP pattern. However, the HetSigma approach presents some interesting features which justify its presentation in this specific subsection.

In the HetSigma approach both the arrival and service processes are modulated in continuous-time by a single infinitesimal generator $\mathbf{Q}_{as}$, with m modulating phase states. This assumption includes as a particular case the situation where the arrival and service processes are modulated individually by infinitesimal generators $\mathbf{Q}_a$ and $\mathbf{Q}_s$ with m_a and m_s phases, respectively. This independent modulation case can be converted into a joint modulation by taking $\mathbf{Q}_{as} = \mathbf{Q}_a \oplus \mathbf{Q}_s$ and $m = m_a m_s$.

Arrivals, under each modulating phase i, consist of the superposition of K independent CPPs of positive arrivals and an independent CPP of negative arrivals. More concretely, the $K+1$ CPPs are described in terms of *generalized exponential* (GE) distributions, which govern exponential inter-arrival times with batches having geometric size distribution. For example, during phase i, the stream of negative arrivals follows a GE distribution with representation (ρ_i, δ_i), which means that a negative batch arrives to the system after an exponential time of rate ρ_i, and its size is $k \geq 1$ with probability $(1-\delta_i)\delta_i^{k-1}$. On the other hand, the service facility has c heterogeneous servers. Each

server is labeled and has its own independent GE service time with parameters (μ_{in}, ϕ_{in}), for $1 \leq n \leq c$ and $1 \leq i \leq m$.

The model description must be completed with a number of queueing specifications including the first come first scheduled for service discipline, a switching policy guaranteeing that the servers labeled with lowest indexes are those rendering service, a killing policy which removes customers at the end of the queue when a negative arrival takes place, and other necessary specifications which are described in detail in Chakka and Do (2007).

3.4. Markov-additive processes of arrivals

In this subsection, we follow Pacheco and Prabhu (1995) to introduce the class of *Markov-additive processes of arrivals*. First of all, we remark that the acronym MAP is used in the literature both for the Markovian arrival process introduced in Section 2 and for the Markov-additive processes of arrivals. For the sake of clarity, here we shall denote the latter as MAPA.

A MAPA is a Markov process with two components X and J. In general, X is a non-Markovian component called the additive component since increments of X correspond to arrivals. The Markov component J sometimes represents an environment factor. In other applications, the phenomenon under study leads naturally to the pair (X,J).

The state space assumed in Pacheco and Prabhu (1995) is $\mathscr{S} = \mathbb{R}^r \times E$, where E is a discrete set. Moreover, it is also assumed that (X,J) is a continuous-time process. Then, a process $(X,J) = \{(X(t),J(t)); t \geq 0\}$ on $\mathscr{S}$ is a MAPA if

(i) (X,J) is a Markov process.
(ii) For all $s \geq 0$ and $t \geq 0$, the conditional distribution of $(X(t+s) - X(s), J(t+s))$, given $(X(s),J(s))$, depends only on $J(s)$.

The above definition follows the spirit of Çinlar (1972a,b), who assumed a more general space E. It is convenient to extend E including a special state Δ which indicates the termination of the process (X,J). Some interesting properties including closure properties under linear transformations and linear combinations can be investigated. On the other hand, to study the lack of memory property, inter-arrival times, moments of the number of counts and other structural properties, it is convenient to reduce to the state space $\mathscr{S} = \mathbb{N}^r \times E$. In this context, the dynamics of the MAPA comprise three types of transitions: (a) arrivals without change of state in J; (b) changes of state in J without arrivals; and (c) arrivals with change of state in J.

Secondary recording of the MAPA is a mechanism that generates a secondary arrival process from the original arrival process. This mechanism includes interesting features like thinning and marking.

Closely related to the MAPA is the class of MMAPs defined for the case where E is finite; see Subsection 3.2. The BMAP corresponds to the simple case with $r = 1$ and $E = \{1,\ldots,m\}$.

The contribution by Pacheco and Prabhu (1995) is generalized in Breuer (2003) to cover the inhomogeneous case. The inhomogeneous BMAP is defined as a MAPA (X,J) with additive space $\mathbb{N}$, finite phase space $E=\{1,\ldots,m\}$ and time-inhomogenous structure for the generator functions

$$\mathbf{Q}(t)=\begin{pmatrix} \mathbf{D}_0(t) & \mathbf{D}_1(t) & \mathbf{D}_2(t) & \mathbf{D}_3(t) & \cdots \\ & \mathbf{D}_0(t) & \mathbf{D}_1(t) & \mathbf{D}_2(t) & \cdots \\ & & \mathbf{D}_0(t) & \mathbf{D}_1(t) & \cdots \\ & & & \ddots & \ddots \end{pmatrix},$$

where the (i,j)th entry of $\mathbf{D}_k(t)$ can be interpreted as the infinitesimal transition rate of recording k arrivals during the infinitesimal interval $(t,t+dt]$ while changing from phase i to phase j. Likewise, other interpretations for BMAPs can be adapted to the time-inhomogeneous case. For example, the matrix $\mathbf{D}(t)=\sum_{k=0}^{\infty}\mathbf{D}_k(t)$ is a generator for all $t\geq 0$. If the phase process J has a stationary distribution $\boldsymbol{\theta}$, then starting the phase process in this distribution without prior arrivals yields the following expression for the mean number of arrivals until time t:

$$\int_0^t \boldsymbol{\theta}\sum_{k=1}^{\infty} k\mathbf{D}_k(u)\mathbf{e}_m du.$$

Breuer (2003) also generalizes the notion of characteristic sequence slightly in order to define a class of fluid MAPs. In this generalization, the phase space is finite $E=\{1,\ldots,m\}$ and the additive space is given by $[0,\infty)$. Unlike the additive space $\mathbb{N}$ which allows us to arrange the matrices containing arrival rates in a single sequence, an analogue for the additive space $[0,\infty)$ is a characteristic measure $\boldsymbol{\Delta}$ providing an arrival rate matrix for every Borel-measurable subset of $[0,\infty)$. For the homogeneous fluid MAP, the measure $\boldsymbol{\Delta}$ is specified by the matrices $\boldsymbol{\Delta}(x)$, whose (i,j)th elements are given by the corresponding infinitesimal transition rates $q(i;[0,x]\times\{j\})$, for $x\geq 0$ and $1\leq i,j\leq m$. Thus, the matrix $\boldsymbol{\Delta}(x)$ has an analogous meaning as the matrix $\mathbf{D}_k$ for the BMAP. The infinitesimal generator of J is given by $\mathbf{D}=\lim_{x\to\infty}\boldsymbol{\Delta}(x)$. Let $\boldsymbol{\theta}$ be its stationary probability vector. Then,

$$\int_0^{\infty} \boldsymbol{\theta} u d\boldsymbol{\Delta}(u)\mathbf{e}_m t$$

gives the expected number of arrivals until time t, if the process starts without prior arrivals and in phase equilibrium $\boldsymbol{\theta}$. It can be also shown that $\lim_{t\to\infty}X(t)/t=\int_0^{\infty}\boldsymbol{\theta}ud\boldsymbol{\Delta}(u)\mathbf{e}_m$, almost surely for all initial phase distributions.

The concept of BMAP can be even generalized towards a class of time-space processes, called spatial MAPs; see Breuer (2003, Chapters 7-9), and Breuer and Baum (2005, Chapter 14). This generalization addresses three essential points: (a) the phase

state E is allowed to be general; (b) the generator functions of the spatial MAP may depend on time; and (c) arrivals may assume a location in some space.

Based on an underlying MAPA, Sengupta (1989) defines a bivariate Markov process (X,J) with a special structure, which can be seen as a continuous-time and continuous-space version of the Markov chains of $GI/M/1$-type studied by Neuts (1981). The *Sengupta process* yields a notably simplified characterization of the waiting time and the queue length distributions in the $GI/PH/1$ queue. Specifically, the phase space is finite $E=\{1,\ldots,m\}$, and the additive component X is skip-free to the right, takes values in $[0,\infty)$ and increases at a linear rate of 1, if there is no downward jump. Moreover, changes in the state of the process (X,J) may also occur in one of two ways:

(i) If $(X(t),J(t))=(x,i)$, then (X,J) may change its state to somewhere between $(x-u,j)$ and $(x-u+du,j)$ at a rate of $da_{ij}(u)$, for $u\in[0,x)$ and $1\le i,j\le m$.
(ii) If $(X(t),J(t))=(x,i)$, then it may transit from (x,i) to $(0,j)$ at a rate of $b_{ij}(x)$, for $x>0$ and $1\le i,j\le m$.

The level-dependent rates $a_{ij}(x)$ and $b_{ij}(x)$ satisfy the condition

$$\sum_{j=1}^{m}(a_{ij}(x)+b_{ij}(x))=-d_i,\quad x>0,\ 1\le i\le m,$$

where $-d_i$ is the rate at which the next state change can occur from the initial state (x,i). This equality clearly implies that the probability that the additive component X takes a downward jump of $u\in[0,x)$ units from x, given that a downward jump occurs, does not depend on the initial level x.

For a related work, we also refer to the bivariate Markov process (X,J) analyzed by Tweedie (1982), where the additive component X takes values in $\mathbb{N}$ and the Markov component J takes values on a general set such as an interval of the real line.

3.5. The BSDE approach

The rationale for using Markovian arrival processes and PH distributions has been already discussed in Section 2. However, the price to be paid frequently in practice is a significant burden on computational time and memory needed due to the high dimensionality of the resulting block-structured Markov chains. The complexity of the underlying stochastic models increases drastically in non-homogeneous settings, where an arbitrary, even infinite number of MAPs and/or PH distributions could be involved. The BSDE approach provides a versatile tool to deal with a non-exponential model with correlated flows, but keeping the dimensionality of the block-structured Markov chain tractable.

In the BSDE approach, we are concerned with a multidimensional continuous-time Markov chain $(\mathbf{X},\mathbf{Y})=\{(X_1(t),\ldots,X_k(t),Y_1(t),\ldots,Y_l(t));t\ge 0\}$. We assume

that $(\mathbf{X},\mathbf{Y})$ is regular and time-homogeneous; in applications, it is often assumed to be irreducible. The sub-vector $\mathbf{X}(t)=(X_1(t),\ldots,X_k(t))$ provides a k-dimensional description of the fundamental aspects of the system state at time t. On the other hand, the sub-vector $\mathbf{Y}(t)=(Y_1(t),\ldots,Y_l(t))$ is a l-dimensional phase vector which completes the Markovian system description. The state space of $(\mathbf{X},\mathbf{Y})$ is a discrete set $\mathscr{S}_{(\mathbf{X},\mathbf{Y})}$ with $(k+l)$-dimensional elements.

The sojourn time $E_{(\mathbf{x},\mathbf{y})}$ that the Markov chain remains in the state $(\mathbf{x},\mathbf{y})$ is exponentially distributed with rate $\lambda_{(\mathbf{x},\mathbf{y})}$. For a given state $(\mathbf{x},\mathbf{y})$, the p-dimensional random vector $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}=(N_1,\ldots,N_p)|_{(\mathbf{x},\mathbf{y})}$ counts the events taking place when $E_{(\mathbf{x},\mathbf{y})}$ expires. The case when no event is observed is denoted by $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}=\mathbf{0}_p$, whereas the occurrence of an event of type s is associated with $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}=n\mathbf{e}_p(s)$, where $n\in\mathbb{Z}-\{0\}$. For example, $n>1$ denotes a multiple positive jump, $n=-1$ represents a negative jump, etc.

The fundamental state $\mathbf{x}$ is updated in the light of the observed value of $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}$. More concretely, we assume that the resulting fundamental state $\mathbf{x}'$ is of the form $\mathbf{x}'=f\left(\mathbf{x},\mathbf{N}|_{(\mathbf{x},\mathbf{y})}\right)$, where the fundamental state function f has to be specified for each particular Markov chain $(\mathbf{X},\mathbf{Y})$. We notice that $\mathbf{x}'=\mathbf{x}$ if $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}=\mathbf{0}_p$.

It should be noted that the case $\mathbf{N}|_{(\mathbf{x},\mathbf{y})}=\mathbf{0}_p$ implies that the phase state $\mathbf{y}$ jumps to a new state $\mathbf{y}'\neq\mathbf{y}$. In contrast, the existence of proper events may or not be accompanied by a phase change.

The kernel $\{\mathbf{P}_{\mathbf{x}}^{\mathbf{n}};(\mathbf{x},\mathbf{n})\in\mathscr{S}_{(\mathbf{X},\mathbf{N})}\}$ completes the specification of the BSDE approach. The elements $p_{\mathbf{x}}^{\mathbf{n}}(\mathbf{y};\mathbf{y}')$ of the matrix $\mathbf{P}_{\mathbf{x}}^{\mathbf{n}}$ record the probabilities of generating the event $\mathbf{n}$ and a transition from phase $\mathbf{y}$ to phase $\mathbf{y}'$, given that the system state was $(\mathbf{x},\mathbf{y})$ just before $E_{(\mathbf{x},\mathbf{y})}$ expires. Since $E_{(\mathbf{x},\mathbf{y})}$ is a sojourn time, we notice that $p_{\mathbf{x}}^{\mathbf{0}_p}(\mathbf{y};\mathbf{y})=0$.

Finally, the infinitesimal generator $\mathbf{Q}=(q_{(\mathbf{x},\mathbf{y})(\mathbf{x}',\mathbf{y}')})$ of the Markov chain $(\mathbf{X},\mathbf{Y})$ is given by

$$q_{(\mathbf{x},\mathbf{y})(\mathbf{x}',\mathbf{y}')}=\begin{cases}-\lambda_{(\mathbf{x},\mathbf{y})}, & \text{if } (\mathbf{x}',\mathbf{y}')=(\mathbf{x},\mathbf{y}),\\ \lambda_{(\mathbf{x},\mathbf{y})}p_{\mathbf{x}}^{\mathbf{n}}(\mathbf{y};\mathbf{y}'), & \text{if } \mathbf{x}'=f\left(\mathbf{x},\mathbf{N}|_{(\mathbf{x},\mathbf{y})}\right),\\ 0, & \text{otherwise.}\end{cases}$$

If it is desired, then the BSDE approach can be used to construct only a part of the stochastic model. In fact, the BMAP can be readily obtained as a particular case of the BSDE approach; see Artalejo and Gómez-Corral (2010, Example 2.1). The BSDE approach can be easily adapted to the discrete-time setting. Indeed, the above BSDE construction is inspired in a similar discrete mechanism, called discrete block state-dependent arrival distribution, which was introduced in Artalejo and Li (2010) to generate the arrival input of a certain discrete-time queue.

4. Application of the BSDE approach to epidemic models

In this section, we show how the BSDE approach presented in Subsection 3.5 can be used to extend many stochastic systems that use Markov chains to model a biological population. More concretely, we consider the *state-dependent susceptible-infected-susceptible* (SD-SIS) epidemic model which generalizes the scalar SIS model allowing non-exponential infection and recovery times, as well as the existence of correlation. Once the SD-SIS model is constructed, we focus in Subsection 4.2 on the time until the extinction. In Subsection 4.3, the counterpart of the coefficient of correlation between inter-arrival times in the BMAP (see Subsection 2.3.2) is introduced.

4.1. Construction of the SD-SIS model

Firstly, we recall the scalar SIS model (see also Allen (2003)). Consider a closed population of size K. At time t, the population consists of $I(t)$ infected individuals and $S(t) = K - I(t)$ susceptible individuals. In this context, the process $\{I(t); t \geq 0\}$ is assumed to be a birth-and-death process on the state space $\{0, 1, \ldots, K\}$. Let β and γ denote the contact and recovery rates, respectively. Then, the birth rates are defined by $\lambda_i = \beta i(K-i)/K$, for $0 \leq i \leq K$. These rates correspond to transitions occurring when a susceptible individual becomes infected in agreement with the current contacts between $I(t)$ and $S(t)$. On the other hand, the death rates $\mu_i = \gamma i$, for $1 \leq i \leq K$, are associated with the recovery of infected individuals.

The construction of the SD-SIS model is based on a BSDE approach with $k = 1$ and $l = p = 2$. The fundamental state $\mathbf{x} = i$ represents the number of infected individuals, whereas the phase state $\mathbf{y} = (m, n)$ consists of the infection and recovery phases in process at time t. The state space $\mathscr{S}_{(\mathbf{X},\mathbf{Y})}$ is given by

$$\mathscr{S}_{(\mathbf{X},\mathbf{Y})} = \{\overline{0}\} \cup \{(i,m,n); 1 \leq i \leq K, 1 \leq m \leq M, 1 \leq n \leq N\}.$$

We notice that the epidemic ends as soon as there are no infected individuals in the population. Thus, we consider an absorbing macrostate $\overline{0}$ with rate $\lambda_{\overline{0}} = 0$. The individuals do not develop immunity after they recover. As a result, the Markov chain $(\mathbf{X}, \mathbf{Y})$ is reducible and the absorption occurs in a finite time with probability one. The events are associated with infections (i.e., single positive jumps) and recoveries (i.e., single negative jumps). It means that the SD-SIS model can be viewed as a particular case of a finite *state-dependent quasi-birth-and-death* (SD-QBD) process; see Artalejo and Gómez-Corral (2010, Section 3).

Then, the infinitesimal generator $\mathbf{Q}$ of the SD-SIS model has the following non-homogeneous block-tridiagonal structure:

$$\mathbf{Q} = \begin{pmatrix} 0 & \mathbf{0}_g & & & & & \\ \mathbf{q}_{10} & \mathbf{Q}_{11} & \mathbf{Q}_{12} & & & & \\ & \mathbf{Q}_{21} & \mathbf{Q}_{22} & \mathbf{Q}_{23} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \mathbf{Q}_{K-1,K-2} & \mathbf{Q}_{K-1,K-1} & \mathbf{Q}_{K-1,K} \\ & & & & \mathbf{Q}_{K,K-1} & \mathbf{Q}_{KK} \end{pmatrix},$$

where the blocks $\mathbf{Q}_{ii'}$ are square matrices of dimension $g = MN$, for $1 \leq i, i' \leq K$. The column vector $\mathbf{q}_{10}$ describes the motion from states $(1, m, n)$ to the absorbing state 0, for $1 \leq m \leq M$ and $1 \leq n \leq N$.

For the derivation of the blocks $\mathbf{Q}_{ii'}$, we need to introduce families of rate matrices $\{\overline{\mathbf{A}}_i^k; 1 \leq i \leq K-1\}$ and $\{\overline{\mathbf{D}}_i^k; 1 \leq i \leq K\}$, for $k \in \{0,1\}$. The elements $\vec{a}_i^k(m;m')$ are defined by

$$\begin{aligned} \vec{a}_i^0(m;m) &= -\lambda_{(i,m)}^A, \\ \vec{a}_i^0(m;m') &= \lambda_{(i,m)}^A a_i^0(m;m'), \quad m' \neq m, \\ \vec{a}_i^1(m;m') &= \lambda_{(i,m)}^A a_i^1(m;m'). \end{aligned}$$

We observe that $\lambda_{(i,m)}^A$ denotes the rate of the exponential sojourn time E_{im}^A, which ends either when an infection takes place (with or without phase change) or simply when the infection phase is changed (no arrival case). If $i = K$, then the whole population is infected, so we have $\lambda_{(K,m)}^A = 0$. In contrast, $\lambda_{(i,m)}^A > 0$ for $1 \leq i \leq K-1$. The kernel probabilities $a_i^k(m,m')$ are the probabilities of $k \in \{0,1\}$ infections (i.e., positive jumps) and a transition from phase m to phase m', given that $\mathbf{x} = i$. The description of the rate matrices $\overline{\mathbf{D}}_i^k$ is similar and thus it is omitted. By assuming independence between E_{im}^A and the analogue recovery sojourn time E_{in}^D, we have that $\lambda_{(i,m,n)} = \lambda_{(i,m)}^A + \lambda_{(i,n)}^D > 0$.

Under the above BSDE specifications, we finally obtain the following non-zero blocks

$$\begin{aligned} \mathbf{q}_{10} &= \left(\mathbf{I}_M \otimes \overline{\mathbf{D}}_1^1\right) \mathbf{e}_g, \\ \mathbf{Q}_{i,i-1} &= \mathbf{I}_M \otimes \overline{\mathbf{D}}_i^1, \quad 2 \leq i \leq K, \\ \mathbf{Q}_{ii} &= \overline{\mathbf{A}}_i^0 \oplus \overline{\mathbf{D}}_i^0, \quad 1 \leq i \leq K-1, \\ \mathbf{Q}_{KK} &= \mathbf{I}_M \otimes \overline{\mathbf{D}}_K^0, \\ \mathbf{Q}_{i,i+1} &= \overline{\mathbf{A}}_i^1 \otimes \mathbf{I}_N, \quad 1 \leq i \leq K-1. \end{aligned}$$

We now turn our attention to the dimensionality problem. The objective is to deal with a particularization of the rate matrices such that the formulation remains sufficiently tractable, yet enough versatile for computational purposes. To reach this objective, we consider the choice

$$\overline{\mathbf{A}}_i^0 = \frac{\lambda_i}{\lambda}\mathbf{D}_0^A, \qquad \overline{\mathbf{A}}_i^1 = \frac{\lambda_i}{\lambda}\mathbf{D}_1^A, \quad 1 \le i \le K-1,$$

$$\overline{\mathbf{D}}_i^0 = \frac{\mu_i}{\mu}\mathbf{D}_0^D, \qquad \overline{\mathbf{D}}_i^1 = \frac{\mu_i}{\mu}\mathbf{D}_1^D, \quad 1 \le i \le K,$$

where $(\mathbf{D}_0^A, \mathbf{D}_1^A)$ and $(\mathbf{D}_0^D, \mathbf{D}_1^D)$ denote the characteristic matrices of two auxiliary MAPs of orders M and N, respectively. Their corresponding fundamental rates are λ and μ.

Since λ_i and μ_i are the birth-and-death rates of the scalar SIS model, we obtain a BSDE formulation that, given that the current number of infected individuals equals i, the expectations until the next infection and recovery epochs match the corresponding expected values in the scalar SIS model.

4.2. Extinction in the SD-SIS model

The extinction time quantifies the spread of the epidemic on the population and describes the time until the end of the epidemic process. Thus, the time to extinction is an important measure of the persistence of an infection. There exists a vast literature studying the extinction time of stochastic biological models. In this subsection, we extend the study to the SD-SIS model.

We distinguish between a conditional version of the extinction time given an initial state and an unconditional version properly defined. The conditional extinction time $L_{(i,m,n)}$ is defined as the absorption time in 0, given that the initial state of the SD-SIS model is $(\mathbf{x},\mathbf{y}) = (i,m,n)$. Let $\varphi_{(i,m,n)}(s)$ be its Laplace-Stieltjes transform. The vectors $\boldsymbol{\varphi}_i(s) = (\varphi_{(i,1,1)}(s), \ldots, \varphi_{(i,M,N)}(s))'$, for $1 \le i \le K$, and $\boldsymbol{\varphi}(s) = (\boldsymbol{\varphi}_1(s), \ldots, \boldsymbol{\varphi}_K(s))'$ comprise the Laplace-Stieltjes transforms according to the levels determined by the number of infected individuals.

By introducing an initial distribution $\boldsymbol{\tau}$ on the state space $\mathscr{S}_{(\mathbf{X},\mathbf{Y})}$, we arrive to the unconditional version L of the extinction time. From the general theory for continuous-time Markov chains (see e.g. Kulkarni (1995), and Latouche and Ramaswami (1999)), we know that L follows a PH distribution of order Kg with representation $(\boldsymbol{\tau}, \mathbf{M})$, where $\mathbf{M}$ is the submatrix of $\mathbf{Q}$ corresponding to the set of transient states $\mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\overline{0}\}$.

Since the set $\mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\overline{0}\}$ is irreducible, the existence of the inverse $\mathbf{M}^{-1}$ is guaranteed. We may also observe that the starting point of the density function is given by $f_L(0) = -\boldsymbol{\tau}\mathbf{M}\mathbf{e}_{Kg} = \boldsymbol{\tau}_1\mathbf{q}_{10}$, where $\boldsymbol{\tau}_1$ is the sub-vector of $\boldsymbol{\tau}$ containing the initial probabilities $\tau_{(1,m,n)}$ of the level $i = 1$.

Coming back to the unconditional version, we notice that the vector $\boldsymbol{\varphi}(s)$ satisfies the block-tridiagonal system

$$(\mathbf{M} - s\mathbf{I}_{Kg})\,\boldsymbol{\varphi}(s) = -\begin{pmatrix} \mathbf{q}_{10} \\ \mathbf{0}'_{(K-1)g} \end{pmatrix}.$$

By using Euler and Post-Widder algorithms, we can numerically invert the above expression to get the conditional density functions $f_{L_{(i,m,n)}}(x)$ and, consequently, the unconditional density $f_L(x)$; see Cohen (2007).

Finally, we observe that the conditional moments $m^k_{(i,m,n)} = E[L^k_{(i,m,n)}]$, for $(i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}$ and $k \geq 1$, can be computed from the formula

$$\mathbf{m}^k = k! \left(-\mathbf{M}^{-1}\right)^k \mathbf{e}_{Kg}, \quad k \geq 1,$$

or, alternatively, from the recursive expressions

$$\begin{aligned} \mathbf{m}^0 &= \mathbf{e}_{Kg}, \\ \mathbf{m}^k &= -k\mathbf{M}^{-1}\mathbf{m}^{k-1}, \quad k \geq 1, \end{aligned}$$

where $\mathbf{m}^k$ denotes the column vector of dimension Kg containing the moments $m^k_{(i,m,n)}$ in lexicographic order.

The unconditional time to extinction depends on the initial distribution $\boldsymbol{\tau}$. In epidemiology, it is often known that a certain epidemic has been evolving for a long time and that it has not reached the extinction yet. However, it may be very difficult to know the exact distribution $\boldsymbol{\tau}$. In this case, the use of the quasi-stationary distribution is especially interesting. The starting point is the conditional probabilities

$$u_{(i,m,n)}(t) = P\left((\mathbf{X}(t),\mathbf{Y}(t)) = (i,m,n) | L > t\right) = \frac{p_{(i,m,n)}(t)}{1 - p_{\bar{0}}(t)},$$

for $(i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}$, where $p_{(i,m,n)}(t)$ and $p_{\bar{0}}(t)$ are the transient probabilities of the Markov chain $(\mathbf{X},\mathbf{Y})$.

Suppose that the Markov chain starts with the initial distribution $\tau_{(i,m,n)} = P((\mathbf{X}(0),\mathbf{Y}(0)) = (i,m,n))$, for $(i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}$. If there exists a starting distribution $\tau_{(i,m,n)} = u_{(i,m,n)}$, such that $u_{(i,m,n)}(t) = u_{(i,m,n)}$, for all $t \geq 0$, then $\mathbf{u} = (u_{(i,m,n)})$ is called a quasi-stationary distribution. Moreover, there also exists a limiting interpretation which states that $\lim_{t\to\infty} u_{(i,m,n)}(t) = u_{(i,m,n)}$, independently of the initial distribution.

In our case, the set $\mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}$ is finite and irreducible. Then, the quasi-stationary distribution $\mathbf{u}$ amounts to the left eigenvector associated with the eigenvalue with maximal real part of the matrix $\mathbf{M}$; see Darroch and Seneta (1967). This result gives a method for numerical computation.

In what follows, we set $\boldsymbol{\tau} = \mathbf{u}$ and generalize the existing approach for the study of the extinction time $L_{\mathbf{u}}$ in the scalar SIS model (see Norden (1982)) to the SD-SIS model.

By differentiating $u_{(i,m,n)}(t)$ with respect to t, we obtain

$$u'_{(i,m,n)}(t) = \frac{p'_{(i,m,n)}(t)}{1 - p_{\bar{0}}(t)} + \frac{p_{(i,m,n)}(t)\, p'_{\bar{0}}(t)}{(1 - p_{\bar{0}}(t))^2}, \quad (i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}.$$

By combining the above formula and the Kolmogorov forward equation for the absorbing state $\bar{0}$, we find that

$$u'_{(i,m,n)}(t) = \frac{p'_{(i,m,n)}(t)}{1-p_{\bar{0}}(t)} + \frac{p_{(i,m,n)}(t)}{1-p_{\bar{0}}(t)} \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot) u_{(1,\cdot,n)}(t), \quad (i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\},$$

where $\bar{d}_1^1(n;\cdot) = \sum_{n'=1}^{N} \bar{d}_1^1(n;n')$ and $u_{(1,\cdot,n)}(t) = \sum_{m=1}^{M} u_{(1,m,n)}(t)$, for $1 \le n \le N$.

Now, we appeal to the fact that the initial distribution is $\mathbf{u}$ and we thus put $u'_{(i,m,n)}(t) = 0$. Hence, for each $(i,m,n) \in \mathscr{S}_{(\mathbf{X},\mathbf{Y})} - \{\bar{0}\}$, we get the differential equation

$$\begin{aligned} p'_{(i,m,n)}(t) &= -p_{(i,m,n)}(t) \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot) u_{(1,\cdot,n)}, \\ p_{(i,m,n)}(0) &= u_{(i,m,n)}, \end{aligned}$$

which yields the solution

$$p_{(i,m,n)}(t) = u_{(i,m,n)} \exp\left\{ -t \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot) u_{(1,\cdot,n)} \right\}.$$

Finally, for $p_{\bar{0}}(t)$, we now have $p'_{\bar{0}}(t) = \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot) p_{(1,\cdot,n)}(t)$, with $p'_{\bar{0}}(0) = 0$, so that

$$P(L_{\mathbf{u}} \le t) = p_{\bar{0}}(t) \;=\; 1 - \exp\left\{ -t \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot) u_{(1,\cdot,n)} \right\}, \quad t \ge 0.$$

This establishes that the time to extinction, when the initial distribution is the quasi-stationary distribution, has an exponential distribution with rate $1/E[L_{\mathbf{u}}] = \sum_{n=1}^{N} \bar{d}_1^1(n;\cdot)\, u_{(1,\cdot,n)}$.

The following example illustrates the influence of the characteristic matrices and the correlation in the distribution of $L_{\mathbf{u}}$.

Example 4.1 We consider the following three choices for the characteristic matrices $(\mathbf{D}_0^A, \mathbf{D}_1^A)$ and $(\mathbf{D}_0^D, \mathbf{D}_1^D)$:

(i) *Exponential kernel*. We take $M = N = 1$, $\mathbf{D}_0^A = \mathbf{D}_0^D = -1$ and $\mathbf{D}_1^A = \mathbf{D}_1^D = 1$.
(ii) *Erlang-hyperexponential kernel*. We take $M = 3$, $N = 2$ and

$$\mathbf{D}_0^A = \begin{pmatrix} -3 & 3 & 0 \\ 0 & -3 & 3 \\ 0 & 0 & -3 \end{pmatrix}, \qquad \mathbf{D}_1^A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 0 \end{pmatrix},$$

$$\mathbf{D}_0^D = \begin{pmatrix} -1.9 & 0 \\ 0 & -0.19 \end{pmatrix}, \qquad \mathbf{D}_1^D = \begin{pmatrix} 1.71 & 0.19 \\ 0.171 & 0.019 \end{pmatrix}.$$

(iii) *MAP-MAP kernel*. We take $M = N = 3$ and

$$\mathbf{D}_0^A = \begin{pmatrix} -1.00221 & 1.00221 & 0 \\ 0 & -1.00221 & 0 \\ 0 & 0 & -225.75 \end{pmatrix}, \quad \mathbf{D}_1^A = \begin{pmatrix} 0 & 0 & 0 \\ 0.99219 & 0 & 0.01002 \\ 2.2575 & 0 & 223.4925 \end{pmatrix},$$

$$\mathbf{D}_0^D = \begin{pmatrix} -0.87478 & 0.87478 & 0 \\ 0 & -0.87478 & 0 \\ 0 & 0 & -94.76811 \end{pmatrix}, \quad \mathbf{D}_1^D = \begin{pmatrix} 0 & 0 & 0 \\ 0.78730 & 0 & 0.08748 \\ 7.28985 & 0 & 87.47826 \end{pmatrix}.$$

For the above three scenarios, the fundamental rates associated with infection and recovery characteristic matrices are $\lambda = \mu = 1.0$. We notice that scenarios (i) and (ii) are associated with renewal processes and, on the contrary, scenario (iii) has positive correlated infection and recovery times. The values of the coefficients of correlation are 0.48890 and 0.43482, respectively.

Table 1: *$E[\mathbf{u}]$, $\sigma(\mathbf{u})$ and $E[L_{\mathbf{u}}]$ for three scenarios.*

	Scenario (i)	**Scenario (ii)**	**Scenario (iii)**
$E[\mathbf{u}]$	64.48076	60.04070	38.91698
$\sigma(\mathbf{u})$	11.87236	20.12606	44.42737
$E[L_{\mathbf{u}}]$	2094831.60843	1140.40538	7.75147

For a population size $K = 200$ and the rates $\beta = 1.5$ and $\gamma = 1.0$, we summarize in Table 1 the main statistical descriptors; that is, the mean and the standard deviation of $\mathbf{u}$, and the expected value $E[L_{\mathbf{u}}]$.

In Figure 2, we turn our attention to the probability distribution function $P(L_{\mathbf{u}} \leq t)$. In this case, we deal with scenario (iii) with $K = 200$, $\gamma = 1.0$ and $\beta \in \{0.5, 1.0, 1.5\}$.

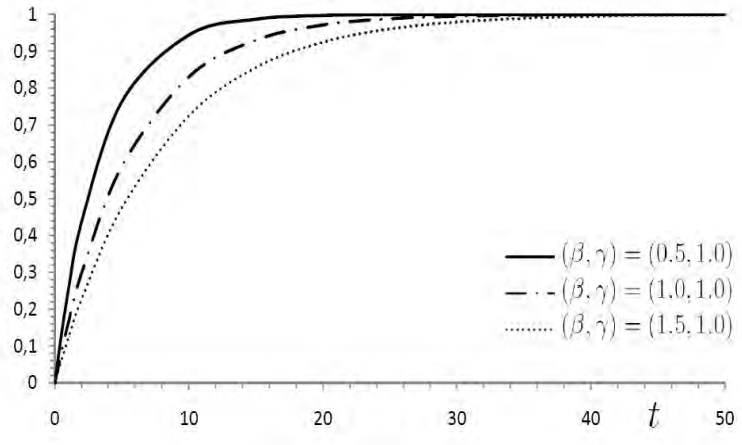


Figure 2: *The probability distribution function $P(L_{\mathbf{u}} \leq t)$.*

In the light of the numerical results, the conclusion is that the influence of the scenario is significant. In other words, the underlying distribution and the correlation are important features which cannot be ignored.

4.3. Correlation between successive events

In this subsection we define a coefficient of correlation between two successive events of the SD-SIS process. Assume that the initial distribution is $\boldsymbol{\tau}$ and denote the first two inter-event intervals as X and Y. To avoid trivialities, we also assume that $K \geq 2$.

First of all, we observe that the one-step transition probability matrix governing the embedded Markov chain at event epochs is given by

$$\mathbf{P} = \begin{pmatrix} 1 & & & & & \\ (-\mathbf{Q}_{11}^{-1})\mathbf{q}_{10}\ \mathbf{0}_{g\times g} & & (-\mathbf{Q}_{11}^{-1})\mathbf{Q}_{12} & & & \\ & \ddots & & \ddots & & \ddots \\ & & (-\mathbf{Q}_{K-1,K-1}^{-1})\mathbf{Q}_{K-1,K-2} & \mathbf{0}_{g\times g} & & (-\mathbf{Q}_{K-1,K-1}^{-1})\mathbf{Q}_{K-1,K} \\ & & & (-\mathbf{Q}_{KK}^{-1})\mathbf{Q}_{K,K-1} & & \mathbf{0}_{g\times g} \end{pmatrix}.$$

To construct a coefficient of correlation, we must guarantee the existence of at least two events before the process reaches its extinction. Thus, if $i = 1$, we correct matrix $\mathbf{P}$ by imposing that the next event is an infection. This modification only affects to the blocks associated with the level $i = 1$ of $\mathbf{Q}$, which are now given by $\mathbf{q}_{10}^c = \mathbf{0}_g'$, $\mathbf{Q}_{12}^c = \mathbf{Q}_{12}$ and $\mathbf{Q}_{11}^c = \mathbf{Q}_{11} + \mathrm{diag}(\mathbf{e}_g(1)\mathbf{q}_{10}, \ldots, \mathbf{e}_g(g)\mathbf{q}_{10})$. As a result, the second row of the corrected matrix $\mathbf{P}^c$ becomes $(\mathbf{0}_g', \mathbf{0}_{g\times g}, (-\mathbf{Q}_{11}^c)^{-1}\mathbf{Q}_{12}, \mathbf{0}_{g\times g}, \ldots)$, while the rest of row blocks does not vary.

In calculating the correlation between X and Y, we shall need the marginal density functions of X and Y, and the joint density function of (X,Y). It is easy to show that they are as follows:

$$f_X(x) = \boldsymbol{\tau}(1)\exp\{\mathbf{Q}_{11}^c x\}(-\mathbf{Q}_{11}^c)\mathbf{e}_g + \sum_{i=2}^{K}\boldsymbol{\tau}(i)\exp\{\mathbf{Q}_{ii}x\}(-\mathbf{Q}_{ii})\mathbf{e}_g, \quad x \geq 0,$$

$$f_Y(y) = \sum_{i=1}^{K}\overline{\boldsymbol{\tau}}(i)\exp\{\mathbf{Q}_{ii}x\}(-\mathbf{Q}_{ii})\mathbf{e}_g, \quad y \geq 0,$$

$$\begin{aligned} f_{(X,Y)}(x,y) &= \boldsymbol{\tau}(1)\exp\{\mathbf{Q}_{11}^c x\}\mathbf{Q}_{12}\exp\{\mathbf{Q}_{22}y\}(-\mathbf{Q}_{22})\mathbf{e}_g \\ &\quad + \sum_{i=2}^{K-1}\boldsymbol{\tau}(i)\exp\{\mathbf{Q}_{ii}x\}\mathbf{Q}_{i,i+1}\exp\{\mathbf{Q}_{i+1,i+1}y\}(-\mathbf{Q}_{i+1,i+1})\mathbf{e}_g \\ &\quad + \sum_{i=2}^{K}\boldsymbol{\tau}(i)\exp\{\mathbf{Q}_{ii}x\}\mathbf{Q}_{i,i-1}\exp\{\mathbf{Q}_{i-1,i-1}y\}(-\mathbf{Q}_{i-1,i-1})\mathbf{e}_g, \quad x \geq 0, y \geq 0, \end{aligned}$$

where the vector $\overline{\boldsymbol{\tau}} = (\overline{\boldsymbol{\tau}}(1), \ldots, \overline{\boldsymbol{\tau}}(K))$ is given by

$$\begin{aligned}
\overline{\boldsymbol{\tau}}(1) &= \boldsymbol{\tau}(2)\left(-\mathbf{Q}_{22}^{-1}\right)\mathbf{Q}_{21},\\
\overline{\boldsymbol{\tau}}(2) &= \boldsymbol{\tau}(1)\left(-\mathbf{Q}_{11}^{c}\right)^{-1}\mathbf{Q}_{12} + \boldsymbol{\tau}(3)\left(-\mathbf{Q}_{33}^{-1}\right)\mathbf{Q}_{32},\\
\overline{\boldsymbol{\tau}}(i) &= \boldsymbol{\tau}(i-1)\left(-\mathbf{Q}_{i-1,i-1}^{-1}\right)\mathbf{Q}_{i-1,i} + (1-\delta_{iK})\boldsymbol{\tau}(i+1)\left(-\mathbf{Q}_{i+1,i+1}^{-1}\right)\mathbf{Q}_{i+1,i}, \quad 3 \le i \le K.
\end{aligned}$$

The vector $\overline{\boldsymbol{\tau}}$ can be readily obtained by noticing that $(\mathbf{0}_g, \overline{\boldsymbol{\tau}}) = (\mathbf{0}_g, \boldsymbol{\tau})\mathbf{P}^c$.

From the density functions, it is straightforward to find the first two moments of X and Y, as well as the cross expectation $E[XY]$. They are given by

$$\begin{aligned}
E[X] &= \boldsymbol{\tau}(1)\left(-\mathbf{Q}_{11}^{c}\right)^{-1}\mathbf{e}_g + \sum_{i=2}^{K}\boldsymbol{\tau}(i)\left(-\mathbf{Q}_{ii}^{-1}\right)\mathbf{e}_g,\\
E[X^2] &= 2\left(\boldsymbol{\tau}(1)\left(-\mathbf{Q}_{11}^{c}\right)^{-2}\mathbf{e}_g + \sum_{i=2}^{K}\boldsymbol{\tau}(i)\left(-\mathbf{Q}_{ii}^{-1}\right)^{2}\mathbf{e}_g\right),\\
E[Y] &= \sum_{i=1}^{K}\overline{\boldsymbol{\tau}}(i)\left(-\mathbf{Q}_{ii}^{-1}\right)\mathbf{e}_g,\\
E[Y^2] &= 2\sum_{i=1}^{K}\overline{\boldsymbol{\tau}}(i)\left(-\mathbf{Q}_{ii}^{-1}\right)^{2}\mathbf{e}_g,\\
E[XY] &= \boldsymbol{\tau}(1)\left(-\mathbf{Q}_{11}^{c}\right)^{-2}\mathbf{Q}_{12}\left(-\mathbf{Q}_{22}^{-1}\right)\mathbf{e}_g\\
&\quad + \sum_{i=2}^{K}\boldsymbol{\tau}(i)\left(-\mathbf{Q}_{ii}^{-1}\right)^{2}\left(\mathbf{Q}_{i,i-1}\left(-\mathbf{Q}_{i-1,i-1}^{-1}\right) + (1-\delta_{iK})\mathbf{Q}_{i,i+1}\left(-\mathbf{Q}_{i+1,i+1}^{-1}\right)\right)\mathbf{e}_g.
\end{aligned}$$

The combination of the above expressions leads to the desired coefficient of correlation

$$\rho(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sqrt{Var(X)Var(Y)}}.$$

The initial distribution can be chosen as $\boldsymbol{\tau} = \mathbf{u}_R$, where $\mathbf{u}_R$ denotes the quasi-stationary distribution of the embedded Markov chain between two regular event epochs, with transition matrix $\mathbf{P}$.

5. Bibliographical notes

Within the list of references we may distinguish between two categories of contributions, depending on whether or not they have been cited throughout the main body of this survey.

Papers and books of the first category have allowed us to review the main aspects of the BMAP and its basic properties, as well as related variants, generalizations and new results in the context of the BSDE approach. The reader has been also addressed to the existing survey papers by Asmussen (2000), Chakravarthy (2001,2010) and Neuts (1992) on the PH distribution and the BMAP and, in a more general setting, to the monographs by Bini *et al.* (2005), Latouche and Ramaswami (1999), Li (2010) and Neuts (1981,1989) which present the main results and algorithms of the matrix-analytic theory.

Regarding to the second category, we associate those papers we do not cite in preceding sections to our desire to present a few selected references dealing with the problem of estimating parameters, multiple types of customers and applications. They are classified as follows:

(i) *Estimation and fitting*

Bodrog *et al.* (2008), Breuer (2002), Breuer and Alfa (2005), Horváth *et al.* (2010), Okamura *et al.* (2009), and Telek and Horváth (2007).

(ii) *Marked arrivals and multiple types of customers*

Alfa *et al.* (2003), He (1996,2000), He and Alfa (2000), Takine and Hasegawa (1994), and Van Houdt and Blondia (2002).

(iii) *Applications*

In queueing and communication systems: Artalejo and Gómez-Corral (2008), Asmussen and Møller (2001), Baek *et al.* (2008), Chakravarthy *et al.* (2006), Choi *et al.* (2004), Daikoku *et al.* (2007), Dudin and Nishimura (1999), He (2001), Kim and Kim (2010), Kim *et al.* (2010), Lambert *et al.* (2006), Li *et al.* (2006), Lucantoni *et al.* (1994), Ost (2001), Shin (2004), Squillante *et al.* (2008), Takine (1999), and Tian and Zhang (2006).

In reliability and maintenance models: Chakravarthy and Gómez-Corral (2009), Frostig and Kenzin (2009), and Montoro-Cazorla and Pérez-Ocón (2008).

In inventory systems: Cheng and Song (2001), He *et al.* (2002), Manuel *et al.* (2007) and Ramaswami (1981).

In risk and insurance problems: Ahn and Badescu (2007), Badescu *et al.* (2007), and Cheung and Landriault (2009).

Since an exhaustive bibliographical work should include several hundreds of papers on the subject in stochastic modelling, we have elaborated the above list only for illustrative purposes.

Acknowledgments

J. R. Artalejo and A. Gómez-Corral were supported by the Government of Spain (Ministry of Science and Innovation) and the European Commission through project MTM2008-01121.

Appendix: Glossary of notation

To begin with, matrices have uppercase letters and vectors lowercase letters. The transpose of $\mathbf{A}$ is written as $\mathbf{A}'$. The matrix $\mathrm{diag}(a_1,\ldots,a_p)$ is the square matrix having elements a_1, ..., a_p along its diagonal and zeros elsewhere.

We denote by $\mathbf{I}_p$ and $\mathbf{0}_{p\times q}$ the identity matrix of order p and the null matrix of dimension $p\times q$, respectively. We let $\mathbf{e}_p$ be the column vector of order p of 1s, and $\mathbf{0}_p$ be the row vector of order p of 0s. The vector $\mathbf{e}_p(j)$ is a column vector of order p such that all entries equal 0, except for the jth one which is equal to 1.

For a square matrix $\mathbf{A}$, the matrix exponential, denoted by $\exp\{\mathbf{A}\}$, is defined by

$$\exp\{\mathbf{A}\}=\sum_{k=0}^{\infty}\frac{1}{k!}\mathbf{A}^k.$$

Consider a matrix $\mathbf{A}=(a_{ij})$ of dimension $p\times q$ and a matrix $\mathbf{B}$ of dimension $r\times s$. The Kronecker product of these matrices, denoted by $\mathbf{A}\otimes\mathbf{B}$, is defined as the structured matrix of dimension $pr\times qs$

$$\mathbf{A}\otimes\mathbf{B}=\begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1q}\mathbf{B}\\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2q}\mathbf{B}\\ \vdots & \vdots & \ddots & \vdots\\ a_{p1}\mathbf{B} & a_{p2}\mathbf{B} & \cdots & a_{pq}\mathbf{B}\end{pmatrix}.$$

Given two square matrices $\mathbf{A}$ and $\mathbf{B}$ of orders p and q, respectively, their Kronecker sum, denoted by $\mathbf{A}\oplus\mathbf{B}$, is defined as the matrix $\mathbf{A}\oplus\mathbf{B}=\mathbf{A}\otimes\mathbf{I}_q+\mathbf{I}_p\otimes\mathbf{B}$.

The Kronecker delta δ_{ij} takes the values 1 if $i=j$, and 0 if $i\neq j$.

References

Ahn, S. and Badescu, A. L. (2007). On the analysis of the Gerber-Shin discounted penalty function for risk processes with Markovian arrivals. *Insurance: Mathematics and Economics*, 41, 234-249.

Alfa, A. S. and Neuts, M. F. (1995). Modelling vehicular traffic using the discrete time Markovian arrival process. *Transportation Science*, 29, 109-117.

Alfa, A. S., Liu, B. and He, Q.-M. (2003). Discrete-time analysis of $MAP/PH/1$ multiclass general pre-emptive priority queue. *Naval Research Logistics*, 50, 662-682.

Allen, L. J. S. (2003). *An Introduction to Stochastic Processes with Applications to Biology*. New Jersey: Prentice Hall.

Artalejo, J. R. and Gómez-Corral, A. (2008). *Retrial Queueing Systems: A Computational Approach*. Berlin: Springer-Verlag.

Artalejo, J. R. and Gómez-Corral, A. (2010). A state-dependent Markov-modulated mechanism for generating events and stochastic models. *Mathematical Methods in the Applied Sciences*, 33, 1342-1349.

Artalejo, J. R. and Li, Q.-L. (2010). Performance analysis of a block-structured discrete-time retrial queue with state-dependent arrivals. *Discrete Event Dynamic Systems*, 20, 325-347.

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

Asmussen, S. and Bladt, M. (1999). Point processes with finite-dimensional conditional probabilities. *Stochastic Processes and their Applications*, 82, 127-142.

Asmussen, S. (2000). Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics*, 27, 193-226.

Asmussen, S. and Møller, J. R. (2001). Calculation of the steady state waiting time distribution in $GI/PH/c$ and $MAP/PH/c$ queues. *Queueing Systems*, 37, 9-29.

Badescu, A. L., Drekic, S. and Landriault, D. (2007). Analysis of a threshold dividend strategy for a MAP risk model. *Scandinavian Actuarial Journal*, 2007, 227-247.

Baek, J. W., Lee, H. W., Lee, S. W. and Ahn, S. (2008). A factorization property for $BMAP/G/1$ vacation queues under variable service speed. *Annals of Operations Research*, 160, 19-29.

Bini, D. A., Latouche, G. and Meini, B. (2005). *Numerical Methods for Structured Markov Chains*. Oxford: Oxford University Press.

Blondia, C. and Casals, O. (1992). Statistical multiplexing of VBR sources: A matrix-analytic approach. *Performance Evaluation*, 16, 5-20.

Bodrog, L., Horváth, A. and Telek, M. (2008). Moment characterization of matrix exponential and Markovian arrival processes. *Annals of Operations Research*, 160, 51-68.

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

Breuer, L. (2003). *From Markov Jump Processes to Spatial Queues*. Dordrecht: Kluwer Academic Publishers.

Breuer, L. and Alfa, A. S. (2005). An EM algorithm for platoon arrival processes in discrete time. *Operations Research Letters*, 33, 535-543.

Breuer, L. and Baum, D. (2005). *An Introduction to Queueing Theory and Matrix-Analytic Methods*. Dordrecht: Springer.

Chakka, R. and Do, T. V. (2007). The $MM\sum_{k=1}^{K} CPP_k/GE/c/L$ G-queue with heterogenous servers: Steady state solution and an application to performance evaluation. *Performance Evaluation*, 64, 191-209.

Chakravarthy, S. R. (2001). The batch Markovian arrival process: A review and future work. In *Advances in Probability & Stochastic Processes*, A. Krishnamoorthy, N. Raju and V. Ramaswami (eds.), Notable Publications, 21-49.

Chakravarthy, S. R., Krishnamoorthy, A. and Joshua, V. C. (2006). Analysis of a multi-server retrial queue with search of customers from the orbit. *Performance Evaluation*, 63, 776-798.

Chakravarthy, S. R. and Gómez-Corral, A. (2009). The influence of delivery times on repairable k-out-of-N systems with spares. *Applied Mathematical Modelling*, 33, 2368-2387.

Chakravarthy, S. R. (2010). Markovian arrival process. In *Wiley Encyclopedia of Operations Research and Management Science*, J. J. Cochran (ed.), John Wiley and Sons, to appear.

Chen, F. and Song, J.-S. (2001). Optimal policies for multiechelon inventory problems with Markov-modulated demand. *Operations Research*, 49, 226-234.

Cheung, E. C. K. and Landriault, D. (2009). Perturbed MAP risk models with dividend barrier strategies. *Journal of Applied Probability*, 46, 521-541.

Ching, W. K. (1997). Markov-modulated Poisson processes for multi-location inventory problems. *International Journal of Production Economics*, 53, 217-223.

Ching, W. K. (2001). *Iterative Methods for Queuing and Manufacturing Systems*. London: Springer-Verlag.

Choi, B. D., Kim, B. and Zhu, D. (2004). $MAP/M/c$ queue with constant impatient time. *Mathematics of Operations Research*, 29, 309-325.

Çinlar, E. (1972a). Markov additive processes. I. *Z. Wahrscheinlichkeitstheory verw. Geb.*, 24, 85-93.

Çinlar, E. (1972b). Markov additive processes. II. *Z. Wahrscheinlichkeitstheory verw. Geb.*, 24, 95-121.

Cohen, A. M. (2007). *Numerical Methods for Laplace Transform Inversion*. New York: Springer.

Daikoku, K., Masuyama, H., Takine, T. and Takahashi, Y. (2007). Algorithmic computation of the transient queue length distribution in the $BMAP/D/c$ queue. *Journal of the Operational Research Society of Japan*, 50, 55-72.

Darroch, J. N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability*, 4, 192-196.

Dudin, A. N. and Nishimura, S. (1999). A $BMAP/SM/1$ queueing system with Markovian arrival input of disasters. *Journal of Applied Probability*, 36, 868-881.

Eckberg, A. E. (1983). Generalized peakedness of teletraffic processes. In *Proceedings of the Tenth International Teletraffic Congress*, Montreal, Canada, paper no. 4, 4B3.

Frostig, E. and Kenzin, M. (2009). Availability of inspected systems subject to shocks – A matrix algorithmic approach. *European Journal of Operational Research*, 193, 168-183.

He, Q.-M. (1996). Queues with marked customers. *Advances in Applied Probability*, 28, 567-587.

He, Q.-M. and Neuts, M. F. (1998). Markov chains with marked transitions. *Stochastic Processes and their Applications*, 74, 37-52.

He, Q.-M. (2000). Quasi-birth-and-death Markov processes with a tree structure and the $MMAP[K]/PH[K]/N/LCFS$ non-preemptive queue. *European Journal of Operational Research*, 120, 641-656.

He, Q.-M. and Alfa, A. S. (2000). Computational analysis of $MMAP[K]/PH[K]/1$ queues with a mixed FCFS and LCFS service discipline. *Naval Research Logistics*, 47, 399-421.

He, Q.-M. (2001). The versatility of $MMAP[K]$ and the $MMAP[K]/G[K]/1$ queue. *Queueing Systems*, 38, 397-418.

He, Q.-M., Jewkes, E. M. and Buzaccot, J. (2002). Optimal and near-optimal inventory control policies for a make-to-order inventory-production system. *European Journal of Operational Research*, 141, 113-132.

He, Q.-M. (2010). Construction of continuous time Markov arrival processes. *Journal of Systems Science and Systems Engineering*, 19, 351-366.

Horváth, A., Horváth, G. and Telek, M. (2010). A joint moments based analysis of networks of $MAP/MAP/1$ queues. *Performance Evaluation*, 67, 759-788.

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.

Kim, C. S., Klimenok, V. I., Mushko, V. and Dudin, A. N. (2010). The $BMAP/PH/N$ retrial queueing system operating in Markovian random environment. *Computers & Operations Research*, 37, 1228-1237.

Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operations Research*, 37, 151-158.

Kulkarni, V. G. (1995). *Modelling and Analysis of Stochastic Systems*. London: Chapman & Hall.

Lambert, J., Van Houdt, B. and Blondia, C. (2006). Queues with correlated service and inter-arrival times and their application to optical buffers. *Stochastic Models*, 22, 233-251.

Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modelling*. Philadelphia: ASA-SIAM.

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.

Li, Q.-L. (2010). *Constructive Computation in Stochastic Models with Applications: The RG-factorizations*. Beijing, Berlin Heidelberg: Tsinghua University Press, Springer-Verlag.

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

Lucantoni, D. M. (1991). New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46.

Lucantoni, D. M. (1993). The $BMAP/G/1$ queue: A tutorial. In *Performance Evaluation of Computer and Communication Systems*, L. Donatiello and R. Nelson (eds.), Lecture Notes in Computer Science, Vol. 729, Springer-Verlag, 330-358.

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

Manuel, P., Sivakumar, B. and Arivarignan, G. (2007). A perishable inventory system with service facilities, MAP arrivals and PH-service times. *Journal of Systems Science and Systems Engineering*, 16, 62-73.

Milne, C. (1982). Transient behaviour of the interrupted Poisson process. *Journal of the Royal Statistical Society. Series B*, 44, 398-405.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2006). Reliability of a system under two types of failures using a Markovian arrival process. *Operations Research Letters*, 34, 525-530.

Montoro-Cazorla, D. and Pérez-Ocón, R. (2008). A maintenance model with failures and inspection following Markovian arrival processes and two repair modes. *European Journal of Operational Research*, 186, 694-707.

Narayana, S. and Neuts, M. F. (1992). The first two moment matrices of the counts for the Markovian arrival process. *Stochastic Models*, 8, 459-477.

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

Neuts, M. F. (1981). *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore: The Johns Hopkins University Press.

Neuts, M. F. (1989). *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. New York: Marcel Dekker, Inc.

Neuts, M. F. (1992). Models based on the Markovian arrival process. *IEICE Transactions on Communications*, E75-B, 1255-1265.

Neuts, M. F., Liu, D. and Narayana, S. (1992). Local poissonification of the Markovian arrival process. *Stochastic Models*, 8, 87-129.

Neuts, M. F. (1993). The burstiness of point processes. *Stochastic Models*, 9, 445-466.

Neuts, M. F. (1995). *Algorithmic Probability: A Collection of Problems*. London: Chapman & Hall.

Neuts, M. F. and Li, J.-M. (1997). An algorithm for the $P(n,t)$ matrices of a continuous BMAP. In *Matrix-analytic Methods in Stochastic Models*, S.R. Chakravarthy and A.S. Alfa (eds.), Lecture Notes in Pure and Applied Mathematics, Vol. 183, Marcel Dekker, 7-19.

Nielsen, B. F., Nilsson, L. A. F., Thygesen, U. H. and Beyer, J. E. (2007). Higher order moments and conditional asymptotics of the batch Markovian arrival process. *Stochastic Models*, 23, 1-26.

Norden, R. H. (1982). On the distribution of the time to extinction in the stochastic logistic population model. *Advances in Applied Probability*, 14, 687-708.

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

Ost, A. (2001). *Performance of Communication Systems: A Model-based Approach with Matrix-geometric Methods.* Berlin: Springer-Verlag.

Pacheco, A. and Prabhu, N. U. (1995). Markov-additive processes of arrivals. In *Advances in Queuing: Theory, Methods and Open Problems*, J.H. Dshalalow (ed.), CRC Press, 167-194.

Ramaswami, V. (1980). The $N/G/1$ queue and its detailed analysis. *Advances in Applied Probability*, 12, 222-261.

Ramaswami, V. (1981). Algorithms for a continuous-review (s,S) inventory system. *Journal of Applied Probability*, 18, 461-472.

Rudemo, M. (1973). Point processes generated by transitions of Markov chains. *Advances in Applied Probability*, 5, 262-286.

Sengupta, B. (1989). Markov processes whose steady state distribution is matrix-exponential with an application to the $GI/PH/1$ queue. *Advances in Applied Probability*, 21, 159-180.

Shin, Y. W. (2004). $BMAP/G/1$ queue with correlated arrivals of customers and disasters. *Operations Research Letters*, 32, 364-373.

Squillante, M. S., Zhang, Y., Sivasubramanian, A. and Gautam, N. (2008). Generalized parallel-server fork-join queues with dynamic task scheduling. *Annals of Operations Research*, 160, 227-255.

Takine, T. and Hasegawa, T. (1994). The workload in the $MAP/G/1$ queue with state-dependent services: Its applications to a queue with preemptive resume priority. *Stochastic Models*, 10, 183-204.

Takine, T. (1999). The nonpreemptive priority $MAP/G/1$ queue. *Operations Research*, 47, 917-927.

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

Tian, N. and Zhang, Z. G. (2006). *Vacation Queueing Models: Theory and Applications*, International Series in Operations Research & Management, Vol. 93. New York: Springer.

Tweedie, R. L. (1982). Operator-geometric stationary distributions for Markov chains, with application to queueing models. *Advances in Applied Probability*, 14, 368-391.

Van Houdt, B. and Blondia, C. (2002). The delay distribution of a type *k* customer in a first-come-first-served $MMAP[K]/PH[K]/1$ queue. *Journal of Applied Probability*, 39, 213-223.

Discussion of "Markovian arrivals in stochastic modelling: a survey and some new results" by Jesús R. Artalejo, Antonio Gómez-Corral and Qi-Ming He

Rafael Pérez Ocón
Departamento de Estadística e Investigación Operativa
Universidad de Granada, España

Matrix-analytic methods (MAMs) have become an important tool for studying complex systems. They preserve the Markovian structure and present the results in a tractable manner. These methods are based in two fundamental elements: the phase-type distributions (PH-distributions) and the Markovian arrival processes (MAPs). Given the potential of these methods, new results and applications arise frequently, and a survey of these methods is very useful from time to time. The paper initiates considering the batch Markovian arrival processes (BMAPs) and describing their properties. The associated counting processes and the descriptors for quantifying the main quantities are given. These processes are introduced in a methodological way, considering examples and particular cases for a better comprehension of how they operate. The application of the methods in queueing, inventories, and reliability is interesting. Variants of the BMAPs that are proven to be useful in applications, the MMAPs and the MAPAs are presented. The BMAPs occupy a central role in the queueing theory, and it is expected that the study and use of these variants will be increasing with time, not only in queueing, but in others domains of application. This part of the paper resumes and illustrates the properties and applications of these classes of processes. The construction of algorithms and computational programs would complete the present paper; it is a challenge for specialists in these topics.

The introduction of block-structured state-dependent event (BSDE) approach for the treatment of stochastic models is an important contribution. Based in the Markovian structure by means of the introduction of phases, this approach allows constructing stochastic models for complex systems. It can be used in the discrete and continuous cases, and some Markovian stochastic models governed by particular MAPs can be deduced from the BSDE approach. The application of the BSDE to the epidemic models illustrates the power of the method, and contributes to consider non-homogeneous stochastic models, involving non-exponential times and the existence of correlation between successive events. The introduction of the non-homogeneity in the MAMs enlarges the possibility of applications that would be very difficult to do following another methodology. The results are complex, but they can be presented in an algorithmic form as a consequence of the MAMs. The incorporation of a methodology and algorithms to elucidate the structure of the BSDE would be useful in the application of this technique for solving problems in different domains of activity.

In the study of stochastic models three are the elements to be considered: modelling, applications, and inference. Modelling and applications must involve methods to be tractable mathematically. The present survey completes and updates previous ones related to modelling and application. Given the complexity of the methods and the speediness of the applications, this is an excellent paper to know the state of the art of the Markovian arrival processes at the present moment.

Thinking of the applications, the paper can be extended in aspects of inference. Essential for the use of MAPs in practice are the numerical algorithms to fit these processes, and the statistical methods for applying to dataset. In the Bibliographical notes in the paper some references about estimation and fitting are given. Related to the fit of phase-type distributions and to the Markov-modulated Poisson process (MMPP), the paper of Asmussen (1997) shows that the EM algorithm can be successfully applied to maximum-likelihood (ML) estimates in Markov models, even in the case of incomplete data, and computational programs for the treatment of the data are constructed and their properties commented. The paper of Asmussen alludes to the previous one of Ryden (1996), where the problem on identifiability and the order of the involved Markov processes in these two particular cases is presented. An area for future research is the inclusion of problems related to the identifiability of general MAPs into the matrix-analytic methods. This will allow to extend the use of MAPs and solve problems that cannot be addressed with the actual knowledge of the inference about these processes.

Asmussen (1997). Phase-type Distributions and Related Point Processes: Fitting and Recent Advances. In: *Matrix-analytic methods in stochastic models*. Chakravarthy, S. R. and Alfa, A. S. (Eds). Marcel Dekker, New York, 137-149.

Ryden (1996). On identifiability and order of continuous-time aggregated Markov chains, Makov-modulated Poisson processes, and phase-type distributions. *Journal of Applied Probability*, 33, 640-653.

Miklos Telek
Budapest University of Technology and Economics
Department of Telecommunications

The paper mainly presents a survey of Markovian arrival process models. It is always hard to decide the level of knowledge of the aimed audience of a paper or a scientific presentation. I think that the goal of a survey paper should be to introduce the main concepts of a field to those who are not that familiar with them yet. Assuming it is the goal of this paper I recommend to be more detailed and precise with the introduction of the applied concepts, a list of explicit points for considerations are forwarded to the authors.

Section 2 starts with the introduction of BMAPs. It is based on a short summary of PH distributions. I would recommend to unify all PH distribution related content into this part.

In a paper like this I prefer derivations starting from a limited number of initial expression than list of final expressions! The majority of the presented complex expressions on MAP properties can be obtained in simple steps from the joint density functions. I recommend at least indicating how to obtain the presented properties (e.g. on page 113).

The relation of structured Markov processes, like quasi birth death processes (QBD), and those generalization of MAPs which account for the arrival and departure of customers (HetSigma, BSDE) is not expressed in the papers. These models can be viewed as queueing systems resulting structured Markov processes. As a consequence efficient computational methods developed for the analysis of structured Markov processes can be applied for the analysis of these arrival processes. A discussion about this relation would further enhance the paper.

The paper introduces the basic theory of various Markovian arrival processes and presents several examples to indicate the wide spread applicability of this versatile set of models. To make this picture complete it would be interesting to add the basic limitations of these models which needs to be considered when applying them in practice.

Some of these limitations are inherited from PH distributions. The most well know one is about the coefficient of variation of the inter-event time distribution which is greater or equal to $1/n$ when the state space of the modulating process is composed by n state. An other typical feature of these models is the exponential asymptotic decay. It holds for a lot of properties like inter-event time distribution, autocorrelation, lag correlation. Beyond these two most well-known ones a set of further practical limitations are published recently. A summary of these limits would be a nice contribution of the manuscript.

Consequently, real systems with quasi deterministic inter-event times or strange decay behaviour or any other property in conflict with the limits of these models cannot be closely modelled with Markovian arrival models. But fortunately also in these cases, in accordance with the denseness property (Section 2.2.4), a computational complexity – accuracy trade-off can be found by increasing the size of the Markovian model.

Yiqiang Q. Zhao
School of Mathematics and Statistics
Carleton University, Canada

First, I would like to congratulate the authors on this excellent comprehensive review on BMAP. This review paper provides readers with easy access to all the important aspects of the BMAP, from its definition to its basic properties; from its history to its extensions; from theoretical aspects to applications.

Applying BMAP in modelling is popular not only because it is a natural generalization of the Poisson process and captures correlations between arrivals, but also because of the more important fact that the use of BMAPs in modelling often leads to a matrix-structured formalism, to which the powerful matrix-analytic method can be applied.

The variants and generalizations touched on in the review paper have been well chosen by the authors, as they also lead to matrix formulations for which analysis can be carried out in terms of matrix-analytic methods. The contents of Section 4 are interesting, though structurally this section seems sidetracked from the main focus of the review. The variants and generalizations of BMAPs could have also gone in a few different directions. One of such alternatives is a comparison, of modelling properties, of the arrival models discussed in the review paper and other commonly seen arrivals, such as arrivals with long-range dependence, Gaussian queues, periodic arrivals and possibly others.

Markov additive processes deserve special attention among all generalizations of BMAPs. The reason for this goes back to the core of the matrix-analytic method. The quasi-birth-and-death (QBD) process is considered an excellent example for explicitly demonstrating some of the key techniques in the core of the matrix-analytic method, such as duality, probabilistic measures under taboo or censoring technique. A comprehensive summary of QBD processes can be found in Latouche and Ramaswami (1999). These techniques, together with Wiener-Hopf factorizations including RG-factorizations and block-form generating functions (or exponential change of matrix (measure)), lead to a concise treatment of the more general matrix-structured paradigm, the GI/G/1 type of matrices in parallel to that for the QBD process, for example, see Zhao, Li and Braun (1998, 2003). The sequence of the non-boundary matrices in the GI/G/1 paradigm leads to a Markov additive process with finitely many background states. It is of interest to notice that the above mentioned techniques are in fact key general tools and methods for queues in applied probability, for example, see Asmussen (2003).

Standard matrix-analytic methods deal with matrices of finite size, like BMAPs, since the method, in both theoretical and computational aspects, relies on properties of

finite dimensional linear spaces or finite matrices. Attempts to generalize finite matrices to infinite ones have a long history dating back to the early 80s, including Tweedie (1982), Ramaswami and Taylor (1996), and Shi, Guo and Liu (1996), among others. Although basic formalizations stand valid for models with infinite matrices, such as the operator-geometric solution and generalized phase type distributions described by an absorbing Markov chain with infinitely many states, there are two main challenges when finite matrices are extended to infinite ones: (1) many key properties from linear algebra are no longer valid for infinite matrices and instead infinite dimensional linear operators now play a key role; and (2) additional non-trivial efforts should be made to address computational issues of the R- and G-measures since they are no longer finite matrices. Recently, analysis of exact tail asymptotics in the stationary probability distribution for a model whose non-boundary matrices defines an additive process with an infinite background space has been a central topic in terms of (extended) matrix-analytic methods. Tail asymptotics can lead to various performance bounds and accurate approximations. The core of extended matrix-analytic methods consists of the same general tools used in the applied probability mentioned above, such as limit theorems for Markov renewal processes, censoring, RG-factorizations, duality, exponential change of matrix. These tools and properties of Markov additive processes are the key for the success of expanding matrix-analytic methods. References in this direction include Takahashi, Fujimoto and Makimoto (2001), Haque (2003), Kroese, Scheinhardt and Taylor (2004), Miyazawa (2004), Miyazawa and Zhao (2004), and He, Li, and Zhao (2009), among others.

Finally, it was a great pleasure for me to be invited as a discussant for this interesting review paper.

References

Asmussen, S. (2003). *Applied Probability and Queues*, 2nd edition. Springer.

Haque, L. (2003). *Tail behaviour for stationary distributions for two-dimensional stochastic models*. Ph.D. Thesis, Carleton University, Ottawa, ON, Canada.

He, Q., Li, H. and Zhao, Y.Q. (2009). Light-tailed behaviour in QBD process with countably many phases. *Stochastic Models*, 25, 50-75.

Kroese, D.P., Scheinhardt, W.R.W. and Taylor, P.G. (2004). Spectral properties of the tandem Jackson network, seen as a quasi-birth-and-death process. *Annals of Applied Probability*, 14(4), 2057-2089.

Miyazawa, M. (2004). The Markov renewal approach to M/G/1 type queues with countably many background states. *Queueing Systems*, 46, 177-196.

Miyazawa, M. and Zhao, Y.Q. (2004). The stationary tail asymptotics in the GI/G/1 type queue with countably many background states. *Advances in Applied Probability*, 36(4), 1231-1251.

Ramaswami, V. and Taylor, P.G. (1996). Some Properties of the Rate Operators in level dependent Quasi Birth and Death Processes with a Countable Number of Phases. *Stochastic Models*, 12(1), 143-164.

Shi, D.H., Guo J. and Liu, L. (1996). SPH-distributions and the rectangle iterative algorithm. *Matrix-Analytic Methods in Stochastic Models*, in S. Chakravarthy and A.S. Alfa (eds). New York: Marcel Dekker, 207-224.

Takahashi, Y., Fujimoto, K. and Makimoto, N. (2001). Geometric decay of the steady-state probabilities in a quasi-birth-and-death process with a countable number of phases. *Stochastic Models*, 17(1), 1-24.

Zhao, Y.Q., Li, W. and Braun, W.J. (1998). Infinite block-structured transition matrices and their properties. *Advances in Applied Probability*, 30, 365-384.

Zhao, Y.Q., Li, W. and Braun, W.J. (2003). Censoring, factorizations, and spectral analysis for transition matrices with block-repeating entries. *Methodology and Computing in Applied Probability*, **5**, 35-58.

Rejoinder

First of all, we would like to thank the three invited discussants for the time spent commenting on our paper. We appreciate their constructive and insightful comments, which have made valuable contributions to the understanding of various interesting problems.

We now briefly respond to some of their comments.

Comments from Prof. R. Pérez-Ocón

Prof. Pérez-Ocón comments on the important role played by the matrix-analytic formalism and the Markovian arrival processes in stochastic modelling. We thank the discussant for his positive and kind remarks on the recently introduced BSDE approach. At a first glance, the BSDE approach and the matrix-analytic methods present common elements; e.g. structured Markov chains, phase method. Although the BSDE approach is closely related to the methods developed for structured Markov chains, the aim of the BSDE approach is to reduce the cost caused from an excessive dimensionality in the matrix representation, which frequently occurs in non-homogenous settings where an arbitrary number of MAPs and/or PH distributions are simultaneously involved. In this sense, the BSDE approach goes beyond the commonly used matrix-analytic methods. Thus, we completely agree with the remarks of the discussant about the need of developing methodological and algorithmic tools for practical use of the BSDE approach. In particular, efforts leading to a suitable treatment of the positive recurrence of infinite structured non-homogeneous Markov chains would be welcome.

Other relevant points commented by the discussant are the fitting and inference aspects. We touched these matters only in the bibliographical notes, where some selected references were given. We are happy that the discussant is adding basic references that will assist readers who are interested in pursuing this subject further.

Comments from Prof. M. Telek

Prof. Telek pointed out in a separate communication a number of helpful comments to improve the paper presentation. These comments have been partially taken into account. We have also incorporated some additional citations in the text, which should be helpful for those readers desirous of knowing how to derive the presented properties.

In the opinion of the discussant, the HetSigma and the BSDE approaches can be viewed as queueing systems resulting in structured Markov processes. Regarding to the HetSigma approach, Chakka and Do (2007) clearly assert that transitions from a level to any other level are possible. Therefore, the matrix structure is general and the standard matrix-analytic methods cannot be used directly. We stress that our interest in the HetSigma approach comes from the fact that both the arrival and the service processes are modulated by the same Markov process. On the other hand, the BSDE approach is intended to construct either a specific part (i.e., the arrival process) or a whole stochastic model in state-dependent frameworks where neither a well-posed matrix structure or the reducibility of the resulting Markov chain are assumed. In this setting, it is our opinion that the possibility of using the classical matrix-analytic tools is limited. Further methodological and computational efforts are definitively needed, as it was mentioned by Prof. Pérez-Ocón.

The discussant accurately points out some limitations of the PH distribution and consequently of the BMAP, whose distribution of inter-arrival times is of PH type; see Subsection 2.3.2 of the paper. This fact leads to a geometrically decaying correlation structure which makes the MAPs less suitable to model certain correlated input processes. Despite of this difficulty, Markovian arrival processes have been also used to model arrivals with long-range dependence whose autocovariance function decays slower than exponentially; see the references given in our reply to Prof. Zhao.

As a general comment, it should be noticed that catching properly some real inputs with time dependence implies to use MAPs of an excessive large order. This important issue connects with the computational cost inherent to the matrix-analytic formalism. Thus, the use of MAPs in practice is limited by the existing fitting methods. The development of good fitting methods for MAPs is a very interesting research topic, which has received a significant attention during the last years. In addition to the references in Section 5 of our paper (see also the comments by Prof. Pérez-Ocón), we now just add one more recent paper by Casale et al. (2010). In this paper, the MAP fitting is based on the Kronecker product composition method. The paper provides an exhaustive study that includes a discussion on some fundamental difficulties of MAP fitting. In another related work, Bause et al. (2009) provide an experimental comparison between MAPs and ARMA (*auto regressive moving average*) and ARTA (*auto regressive to anything*) based models. The authors conclude that MAP fitting is most demanding in terms of running time.

Comments from Prof. Y.Q. Zhao

Prof. Zhao points out that the paper did not give a complete survey on the possible variants and generalizations of the BMAP. More concretely, the discussant mentions arrivals with long-range dependence, periodic arrivals and Gaussian queues as other alternative arrival processes. There exists a number of papers (e.g. Andersen and

Nielsen (1998), Casale et al. (2008), and Salvador et al. (2004)) where Markovian arrival processes and, specifically, superpositions of MMPPs are used as a very versatile tool to model variable packet traffic exhibiting long-range dependence. The Hurst parameter introduced by Willinger et al. (1995) is frequently used to measure long-range dependence. Periodic arrivals are related to time-inhomogeneous structures; see Section 3.4 in the paper. We agree that periodic arrivals have interest in modelling communication networks. These arrival inputs include, among others, the periodic Poisson process (see Margolius (2007)) and the periodic BMAP (see Breuer (2003)). Despite of the interest in Gaussian sources and Gaussian queues, it is our opinion that they are not commonly analyzed through those techniques belonging to the core of the matrix-analytic methods. We would recommend the book by Mandjes (2007) to the interested readers.

Other important comments from the discussant are regarding to the relevance of a variety of techniques, such as duality, taboo and censoring, and *RG*-factorizations, in the core of the matrix-analytic methods. The discussant accurately makes observations on these techniques as in fact very general and powerful methods for investigating challenging problems including generalization from finite blocks to Markov chains with infinite blocks. Prof. Zhao provides a set of references that deal with this issue, putting emphasis on tail asymptotic results. These comments are more relevant to matrix-analytic methods in general, rather than Markovian arrival processes. We thank Prof. Zhao for this valuable addition.

Finally, we would like to thank once again the discussants. We sincerely hope that our review paper and their comments will be of interest for the audience of this journal. We also take this opportunity to thank the Editor-in-Chief, M. Guillén, and the Executive Editor, P. Puig, for their kind invitation to write the paper and for organizing the stimulating discussion.

Additional references

Andersen, A. T. and Nielsen, B. F. (1998). A Markovian approach for modelling packet traffic with long-range dependence. *IEEE Journal on Selected Areas in Communications*, 16, 719-732.

Bause, F., Buchholz, P. and Kriege, J. (2009). A comparison on Markovian arrival and ARMA/ARTA processes for the modelling of correlated input processes. In *Proceedings of the 2009 Winter Simulation Conference*, M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls (eds.), IEEE Press, 634-645.

Breuer, L. (2003). *From Markov Jump Processes to Spatial Queues*. Dordrecht: Kluwer Academic Publishers.

Casale, G., Mi, N. and Smirni, E. (2008). Versatile models of systems using MAP queueing networks. In *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing (IPDPS 2008)*, IEEE Press, 1-5.

Casale, G., Zhang, E. and Smirni, E. (2010). KPC-Toolbox: Best recipes for automatic trace fitting using Markovian arrival processes. *Performance Evaluation*, 67, 873-896.

Mandjes, M. (2007). *Large Deviations for Gaussian Queues: Modelling Communication Networks*. Chichester: John Wiley & Sons.

Margolius, B. H. (2007). Transient and periodic solution to the time-inhomogeneous quasi-birth death process. *Queueing Systems*, 56, 183-194.

Salvador, P., Pacheco, A. and Valadas, R. (2004). Modelling IP traffic: Joint characterization of packet arrivals and packet sizes using BMAPs. *Computer Networks*, 44, 335-352.

Willinger, W., Taqqu, M. S., Leland, W. E. and Wilson, D. V. (1995). Self-similarity in high-speed packet traffic: Analysis and modelling of Ethernet traffic measurements. *Statistical Science*, 10, 67-85.

SORT 34 (2) July-December 2010, 157-180

On ratio and product methods with certain known population parameters of auxiliary variable in sample surveys

Housila. P. Singh, Ritesh Tailor[1] and Rajesh Tailor

School of Studies in Statistics
Vikram University, Ujjain – 456010, M.P., India

Abstract

This paper proposes two ratio and product-type estimators using transformation based on known minimum and maximum values of auxiliary variable. The biases and mean squared errors of the suggested estimators are obtained under large sample approximation. Conditions are obtained under which the suggested estimators are superior to the conventional unbiased estimator, usual ratio and product estimators of population mean. The superiority of the proposed estimators are also established through some natural population data sets.

MSC: 94A20

Keywords: Study variate, auxiliary variate bias, mean squared error, simple random sampling without replacement

1. Introduction

The use of supplementary information on an auxiliary variable for estimating the finite population mean of the variable under study has played an eminent role in sampling theory and practices. Out of many ratio, product and regression methods of estimation are good illustrations in this context. When the correlation between the study variable y and the auxiliary variable x is positive (high), the ratio method of estimation is employed. On the other hand if this correlation is negative (high), the product method of estimation investigated by Robson (1957) and Murthy (1964), is quite effective.

[1] Lokmanya Tilak Mahavidyalaya, Ujjain-456006, M.P., India
Received: August 2009
Accepted: September 2010

It is a well-established fact that the ratio estimator is most effective when the relation between y and x is straight line through the origin and the variance of y about this line is proportional to x, for instance, see Cochran (1963). In many practical situations, the regression line does not pass through the origin. Also due to stronger intuitive appeal survey statisticians are more inclined towards the use of ratio and product estimators. Keeping these facts in mind several authors including Srivastava (1967, 1983), Reddy (1973,74), Walsh (1970), Gupta (1978), Vos (1980), Naik and Gupta (1991), Mohanty and Sahoo (1995), Sahai and Sahai (1985), Upadhyaya and Singh (1999), Srivenkataramana (1980), Bandyopadhyaya (1980), Mohanty and Das (1971), Srivenkataramana (1978), Sisodia and Dwivedi (1981) and Singh (2003) have suggested various modifications in ratio and product estimators.

Suppose we have population of N identifiable units on which the two variates y and x are defined. For estimating the population mean $\overline{Y} = \sum_{i=1}^{N} y_i \Big/ N$ of the study variate y, a simple random sample of size n is drawn without replacement. It is assumed that the population mean $\overline{X} = \sum_{i=1}^{N} x_i \Big/ N$ of the auxiliary variate x is known. Then the classical ratio and product estimators of population mean $\overline{Y}$ are respectively defined by

$$\bar{y}_R = \bar{y}(\overline{X}/\bar{x}) \tag{1.1}$$

and

$$\bar{y}_p = \bar{y}(\bar{x}/\overline{X}) \tag{1.2}$$

where $\bar{y} = \sum_{i=1}^{n} y_i/n$ and $\bar{x} = \sum_{i=1}^{n} x_i/n$ are the sample means of variates y and x respectively.

Let x_m and x_M be the minimum and maximum values of a known positive variate x respectively. Using these values (i.e. x_m and x_M), Mohanty and Sahoo (1995) suggested to transform auxiliary variable x to new variables z and u such that

$$z_i = \frac{x_i + x_m}{x_M + x_m} \tag{1.3}$$

and

$$u_i = \frac{x_i + x_M}{x_M + x_m}, \quad i = 1, 2, \ldots, N. \tag{1.4}$$

Using these transformed variables z and u, Mohanty and Sahoo (1995) proposed the following ratio estimators for population mean $\overline{Y}$ as

$$t_{1R} = \bar{y}(\overline{Z}/\bar{z}) \tag{1.5}$$

and

$$t_{2R} = \bar{y}(\overline{U}/\bar{u}), \tag{1.6}$$

where

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i + x_m}{x_M + x_m}\right) = \left(\frac{\bar{x} + x_m}{x_M + x_m}\right) \quad \text{and} \quad \bar{u} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i + x_M}{x_M + x_m}\right) = \left(\frac{\bar{x} + x_M}{x_M + x_m}\right)$$

are sample means of z and u respectively, and

$$\overline{Z} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i + x_m}{x_M + x_m}\right) = \left(\frac{\overline{X} + x_m}{x_M + x_m}\right) \quad \text{and} \quad \overline{U} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i + x_M}{x_M + x_m}\right) = \left(\frac{\overline{X} + x_M}{x_M + x_m}\right)$$

are the population means of z and u respectively.

When the correlation between y and x is negative, the product estimator based on transformed variables z and u are defined by

$$t_{1p} = \bar{y}(\bar{z}/\overline{Z}) \tag{1.7}$$

and

$$t_{2p} = \bar{y}(\bar{u}/\overline{U}) \tag{1.8}$$

It is well known under simple random sampling without replacement (SRSWOR) that the mean squared error (or variance) of $\bar{y}$ is

$$\mathrm{MSE}(\bar{y}) = Var(\bar{y}) = \theta S_y^2 = \theta \overline{Y}^2 C_y^2 \tag{1.9}$$

where $\theta = (N-n)/(nN)$, $C_y = \dfrac{S_y}{\overline{Y}}$: the coefficient of variation of the study variate y.

To the first degree of approximation, the biases and mean squared errors (MSEs) of the ratio-type estimators $\bar{y}_R$, t_{1R}, and t_{2R}, and product-type estimators $\bar{y}_p$, t_{1p} and t_{2p} are respectively given by

$$B(\bar{y}_R) = \theta\, \overline{Y} C_x^2 (1-K) \tag{1.10}$$

$$B(t_{1R}) = \theta\, \overline{Y}\,(C_x^2/C_1)\,\{(1/C_1) - K\} \tag{1.11}$$

$$B(t_{2R}) = \theta\, \overline{Y}\,(C_x^2/C_2)\,\{(1/C_2) - K\} \tag{1.12}$$

$$B(\bar{y}_p) = \theta\, \overline{Y}\, C_x^2 K \tag{1.13}$$

$$B(t_{1p}) = \theta\,\overline{Y}\,(C_x^2/C_1)K \tag{1.14}$$

$$B(t_{2p}) = \theta\overline{Y}\,(C_x^2/C_2)\,K \tag{1.15}$$

$$\mathrm{MSE}(\bar{y}_R) = \theta\,\overline{Y}^2[C_y^2 + C_x^2(1-2K)] \tag{1.16}$$

$$\mathrm{MSE}(t_{1R}) = \theta\,\overline{Y}^2[C_y^2 + (C_x^2/C_1)\ \{(1/C_1) - 2K\}] \tag{1.17}$$

$$\mathrm{MSE}(t_{2R}) = \theta\,\overline{Y}^2[C_y^2 + (C_x^2/C_2)\ \{(1/C_2) - 2K\}] \tag{1.18}$$

$$\mathrm{MSE}(\bar{y}_p) = \theta\,\overline{Y}^2[C_y^2 + C_x^2(1+2K)] \tag{1.19}$$

$$\mathrm{MSE}(t_{1p}) = \theta\,\overline{Y}^2[C_y^2 + (C_x^2/C_1)\,\{(1/C_1) + 2K\}] \tag{1.20}$$

$$\mathrm{MSE}(t_{2p}) = \theta\,\overline{Y}^2[C_y^2 + (C_x^2/C_2)\,\{(1/C_2) + 2K\}] \tag{1.21}$$

where $K = \rho\,C_y/C_x$, $\rho = S_{yx}/(S_x S_y)$ is the correlation coefficient between y and x,

$S_x^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2/(N-1)$, $S_y^2 = \sum_{i=1}^{N}(y_i - \overline{Y})^2/(N-1)$, $S_{xy} = \sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})/(N-1)$,

$C_1 = \left(1 + \frac{x_m}{\overline{X}}\right)$, $C_2 = \left(1 + \frac{x_M}{\overline{X}}\right)$ and $C_x = \frac{S_x}{\overline{X}}$: the coefficient of variation of the auxiliary variate x.

It is to be noted that the transformations (1.3) and (1.4) depend on both maximum (x_M) and minimum (x_m) values but the estimators $t_{1R}(t_{1P})$ and $t_{2R}(t_{2P})$ generated through these transformations depend only on maximum value (x_M) and minimum value (x_m) respectively. For instance,

$$\begin{aligned} t_{1R} &= \bar{y}\frac{\overline{Z}}{\bar{z}} \\ &= \bar{y}\frac{(\overline{X} + x_m)/(x_M + x_m)}{(\bar{x} + x_m)/(x_M + x_m)} \\ &= \bar{y}\frac{(\overline{X} + x_m)}{(\bar{x} + x_m)} \end{aligned} \tag{1.22}$$

In similar fashion it can be shown that the estimators t_{1P} and $(t_{2R,}t_{2P,})$ depend only on x_m and x_M respectively.

Expressions (1.22)–(1.25) motivated authors to investigate some transformations which make use of both maximum value (x_M) and minimum value (x_m) and hence using such transformations the constructed estimators should also depend on x_M and x_m. Some ratio- and product-type estimators of population mean $\overline{Y}$ have been suggested and their properties are studied. Numerical illustrations are given in support of the present study.

2. The suggested transformations and estimators

Let x_m and x_M be the minimum and maximum values of a known positive variate x respectively. Using x_m and x_M, it is suggested to transform the auxiliary variable x to new variables 'a' and 'b' such that

$$a_i = x_M x_i + x_m^2 \tag{2.1}$$

and

$$b_i = (x_M - x_m)x_i + x_m^2 \qquad i = 1,2,\ldots,N. \tag{2.2}$$

Using the transformed variates at (2.1) and (2.2) we define the following ratio-type estimators for population mean $\overline{Y}$ as

$$d_{1R} = \overline{y}\left(\frac{\overline{A}}{\overline{a}}\right) \tag{2.3}$$

$$d_{2R} = \overline{y}\left(\frac{\overline{B}}{\overline{b}}\right) \tag{2.4}$$

and the product-type estimators for $\overline{Y}$ as

$$d_{1p} = \overline{y}\left(\frac{\overline{a}}{\overline{A}}\right) \tag{2.5}$$

and

$$d_{2p} = \overline{y}\left(\frac{\overline{b}}{\overline{B}}\right) \tag{2.6}$$

where

$$\overline{a} = \sum_{i=1}^{n} a_i/n = x_M\overline{x} + x_m^2 \quad \text{and} \quad \overline{b} = \sum_{i=1}^{n} b_i/n = (x_M - x_m)\overline{x} + x_m^2$$

are the sample means of 'a' and 'b' respectively and

$$\overline{A} = \sum_{i=1}^{N} a_i/N = x_M \overline{X} + x_m^2 \quad \text{and} \quad \overline{B} = \sum_{i=1}^{N} b_i/N = (x_M - x_m)\overline{X} + x_m^2$$

are the population means of 'a' and 'b' respectively.

2.1. Biases and variances of ratio-type estimators d_{1R} and d_{2R}

To obtain the biases and variances of d_{1R} and d_{2R}, we write

$$\bar{y} = \overline{Y}(1+e_0)$$

$$\bar{x} = \overline{X}(1+e_1)$$

such that

$$E(e_0) = E(e_1) = 0$$

and

$$\left.\begin{aligned} E(e_0^2) &= \theta C_y^2 \\ E(e_1^2) &= \theta C_x^2 \\ E(e_0 e_1) &= \theta K C_x^2 \end{aligned}\right\} \tag{2.7}$$

Expressing $d_{1R}$and d_{2R} in terms of e's we have

$$\begin{aligned} d_{1R} &= \overline{Y}(1+e_0)\frac{\overline{A}}{\left\{x_M \overline{X}(1+e_1) + x_m^2\right\}} \\ &= \overline{Y}(1+e_0)\frac{\overline{A}}{\left\{x_M \overline{X} + x_m^2 + x_M \overline{X} e_1\right\}} \\ &= \overline{Y}(1+e_0)\frac{\overline{A}}{\left\{\overline{A} + x_M \overline{X} e_1\right\}} \\ &= \overline{Y}(1+e_0)\left(1+\lambda_{(1)} e_1\right)^{-1} \end{aligned} \tag{2.8}$$

$$\begin{aligned} d_{2R} &= \overline{Y}(1+e_0)\frac{\overline{B}}{\left\{(x_M - x_m)\overline{X}(1+e_1) + x_m^2\right\}} \\ &= \overline{Y}(1+e_0)\frac{\overline{B}}{\left\{(x_M - x_m)\overline{X} + x_m^2 + (x_M - x_m)\overline{X} e_1\right\}} \end{aligned}$$

$$= \overline{Y}(1+e_0)\frac{\overline{B}}{\left\{\overline{B}+(x_M-x_m)\overline{X}\,e_1\right\}}$$

$$= \overline{Y}(1+e_0)\left(1+\lambda_{(2)}e_1\right)^{-1} \tag{2.9}$$

where

$$\lambda_{(1)} = \frac{x_M\overline{X}}{x_M\overline{X}+x_m^2} = \frac{x_M\overline{X}}{\overline{A}} = \frac{(C_2-1)}{(C_2-1)+(C_1-1)^2} \tag{2.10}$$

and

$$\lambda_{(2)} = \frac{(x_M-x_m)\overline{X}}{(x_M-x_m)\overline{X}+x_m^2} = \frac{(x_M-x_m)\overline{X}}{\overline{B}} = \frac{(C_2-C_1)}{(C_2-C_1)+(C_1-1)^2} \tag{2.11}$$

We now assume that $\left|\lambda_{(1)}e_1\right| < 1$ and $\left|\lambda_{(2)}e_2\right| < 1$ so that we may expand $(1+\lambda_{(1)}e_1)^{-1}$ and $(1+\lambda_{(2)}e_1)^{-1}$ as a series in power of $\lambda_{(1)}e_1$ and $\lambda_{(2)}e_1$. Expanding right hand sides of (2.8) and (2.9), multiplying out and retaining terms of e's to the second degree, we obtain

$$t_{1R} \cong \overline{Y}\left(1+e_0-\lambda_{(1)}e_1-\lambda_{(1)}e_1e_0+\lambda_{(1)}^2e_1^2\right)$$

or

$$(t_{1R}-\overline{Y}) = \overline{Y}\left(e_0-\lambda_{(1)}e_1-\lambda_{(1)}e_1e_0+\lambda_{(1)}^2e_1^2\right) \tag{2.12}$$

and

$$t_{2R} \cong \overline{Y}\left(1+e_0-\lambda_{(2)}e_1-\lambda_{(2)}e_1e_0+\lambda_{(2)}^2e_1^2\right)$$

or

$$(t_{2R}-\overline{Y}) = \overline{Y}\left(e_0-\lambda_{(2)}e_1-\lambda_{(2)}e_1e_0+\lambda_{(2)}^2e_1^2\right) \tag{2.13}$$

Taking expectations of both sides of (2.12) and (2.13) and using the results in (2.7) we get the biases of d_{1R} and $d_{2R}$to the first degree of approximation respectively as

$$B(d_{1R}) = \theta\,\overline{Y}C_x^2\lambda_{(1)}(\lambda_{(1)}-K) \tag{2.14}$$

and

$$B(d_{2R}) = \theta\,\overline{Y}C_x^2\lambda_{(2)}(\lambda_{(2)}-K) \tag{2.15}$$

It follows from (2.14) and (2.15) that the biases $B(d_{1R})$ and $B(d_{2R})$ are negligible, if the sample size n is large enough.

Squaring both sides of (2.12) and (2.13) and retaining terms of e's to the second degree we have

$$(d_{1R}-\overline{Y})^2=\overline{Y}^2\left(e_0^2+\lambda_{(1)}^2e_1^2-2\lambda_{(1)}e_0e_1\right) \tag{2.16}$$

and

$$(d_{2R}-\overline{Y})^2=\overline{Y}^2\left(e_0^2+\lambda_{(2)}^2e_1^2-2\lambda_{(2)}e_0e_1\right) \tag{2.17}$$

Taking expectation of both sides of (2.16) and (2.17) and using the results in (2.7), we get the MSEs of d_{1R} and d_{2R} to the first degree of approximation respectively as

$$\text{MSE}(d_{1R})=\theta\overline{Y}^2\left[C_y^2+\lambda_{(1)}C_x^2(\lambda_{(1)}-2K)\right] \tag{2.18}$$

and

$$\text{MSE}=(d_{2R})=\theta\overline{Y}^2\left[C_y^2+\lambda_{(2)}C_x^2(\lambda_{(2)}-2K)\right] \tag{2.19}$$

2.2. Biases and variances of product-type estimators

To obtain the biases and MSEs of d_{1P} and d_{2P}, we express d_{1P} and d_{2P} in terms of e's as

$$\begin{aligned}
d_{1P} &= \overline{Y}(1+e_0)\frac{\left\{x_M\overline{X}(1+e_1)+x_m^2\right\}}{(x_M\overline{X}+x_m^2)}\\
&= \overline{Y}(1+e_0)\left\{1+\frac{x_M\overline{X}\,e_1}{(x_M\overline{X}+x_m^2)}\right\}\\
&= \overline{Y}(1+e_0)(1+\lambda_{(1)}e_1)\\
&= \overline{Y}(1+e_0+\lambda_{(1)}e_1+\lambda_{(1)}e_0e_1)
\end{aligned}$$

or

$$(d_{1P}-\overline{Y})=\overline{Y}(e_0+\lambda_{(1)}\,e_1+\lambda_{(1)}e_0e_1) \tag{2.20}$$

$$\begin{aligned}
d_{2P} &= \overline{Y}(1+e_0)\frac{\left\{(x_M-x_m)\overline{X}(1+e_1)+x_m^2\right\}}{\left\{(x_M-x_m)\overline{X}+x_m^2\right\}}\\
&= \overline{Y}(1+e_0)\left\{1+\frac{(x_M-x_m)\overline{X}\,e_1}{\left\{(x_M-x_m)\overline{X}+x_m^2\right\}}\right\}\\
&= \overline{Y}(1+e_0)(1+\lambda_{(2)}e_1)\\
&= \overline{Y}(1+e_0+\lambda_{(2)}e_1+\lambda_{(2)}e_0e_1)
\end{aligned}$$

or

$$(d_{2P}-\overline{Y})=\overline{Y}(e_0+\lambda_{(2)}e_1+\lambda_{(2)}e_0e_1), \tag{2.21}$$

where $\lambda_{(1)}$ and $\lambda_{(2)}$ are respectively given by (2.10) and (2.11).

Taking expectation of both sides of (2.19) and (2.20) and using the results in (2.7), we get the exact biases of d_{1P} and d_{2P} as

$$B(d_{1P})=\theta\,\overline{Y}\,\lambda_{(1)}KC_x^2 \tag{2.22}$$

and

$$B(d_{2P})=\theta\,\overline{Y}\,\lambda_{(2)}KC_x^2 \tag{2.23}$$

Squaring both sides of (2.20) and (2.21) and retaining terms of e's to the second degree, and then taking expections, we get the MSEs of d_{1P} and d_{2P} respectively as

$$\mathrm{MSE}(d_{1P})=\theta\overline{Y}^2\left[C_y^2+\lambda_{(1)}C_x^2(\lambda_{(1)}+2K)\right] \tag{2.24}$$

and

$$\mathrm{MSE}(d_{2P})=\theta\overline{Y}^2\left[C_y^2+\lambda_{(2)}C_x^2(\lambda_{(2)}+2K)\right] \tag{2.25}$$

3. Comparison of biases

The absolute relative bias (ARB) of an estimator t of the population mean $\overline{Y}$ is defined by

$$\mathrm{ARB}(t)=\left|\frac{B(t)}{\overline{Y}}\right| \tag{3.1}$$

where $B(t)$ stands for bias of the estimator t.

The comparison of absolute relative biases of ratio-type and product-type estimators have been made and the conditions are displayed in Tables 3.1 and 3.2 respectively.

Table 3.1: *Comparison of absolute relative biases of ratio-type estimators.*

Estimator	Absolute Relative Bias of	
	d_{1R} is less than	d_{2R} is than
$\bar{y}_R$	if either $K > \left(1+\lambda_{(1)}\right)$ or $K < \dfrac{(1+\lambda_{(1)}^2)}{(1+\lambda_{(1)})}$	if either $K > \left(1+\lambda_{(2)}\right)$ or $K < \dfrac{(1+\lambda_{(2)}^2)}{(1+\lambda_{(2)})}$
t_{1R}	if $\dfrac{(1+\lambda_{(1)}^2 C_1^2)}{C_1(1+\lambda_{(1)}C_1)} < K < \dfrac{(1+\lambda_{(1)}C_1)}{C_1}$	if either $\dfrac{(1+\lambda_{(2)}^2 C_1^2)}{C_1(1+\lambda_{(2)}C_1)} < K < \dfrac{(1+\lambda_{(2)}C_1)}{C_1}$, $C_1 < \dfrac{1}{2}(1+C_2)$ or $K < \dfrac{(1+\lambda_{(2)}^2 C_1^2)}{C_1(1+\lambda_{(2)}^2 C_1)}$, $C_1 < \dfrac{1}{2}(1+C_2)$ or $K > \dfrac{(1+\lambda_{(2)}C_1)}{C_1}$, $C_1 > \dfrac{1}{2}(1+C_2)$
t_{2R}	if $\dfrac{(1+\lambda_{(1)}^2 C_2^2)}{C_2(1+\lambda_{(1)}C_2)} < K < \dfrac{(1+\lambda_{(1)}C_1)}{C_2}$	if either $\dfrac{(1+\lambda_{(2)}^2 C_2^2)}{C_2(1+C_2\lambda_{(2)})} < K < \dfrac{(1+\lambda_{(2)}C_2)}{C_2}$, $\lambda_{(2)}C_2 > 1$ or $K < \dfrac{(1+\lambda_{(2)}^2 C_2^2)}{C_2(1+\lambda_{(2)}C_2)}$, $\lambda_{(2)}C_2 > 1$ or $K > \dfrac{(1+\lambda_{(2)}C_2)}{C_2}$, $\lambda_{(2)}C_2 < 1$
d_{2R}	if $\dfrac{(\lambda_{(1)}^2+\lambda_{(2)}^2)}{(\lambda_{(1)}+\lambda_{(2)})} < K < (\lambda_{(1)}+\lambda_{(2)})$	—

It can be easily proved that d_{1P} has smaller absolute relative bias (ARB) than the conventional product estimator $\bar{y}_p$ but larger than that of Mohanty and Sahoo's (1995) estimators t_{1p} and t_{2p}. Table 3.2 clearly indicates that the proposed estimator d_{2P} has smaller absolute relative bias than the conventional product estimator $\bar{y}_P$ as the condition $\lambda_{(2)} < 1$ always holds.

Table 3.2: *Comparison of absolute relative biases of product-type estimators.*

Estimator	Absolute Relative Bias of d_{2P} is less than
$\bar{y}_P$	if $\lambda_{(2)} < 1$
t_{1P}	if $\lambda_{(2)} < \frac{1}{C_1}, \quad C_1 > \frac{(1+C_2)}{2}$
t_{2P}	if $\left\|C_1^2 + C_1\,(C_2-3) - C_2\,(C_2-1)+1\right\| > 0$
d_{1P}	if $\lambda_{(2)} < \lambda_{(1)}$

4. Efficiency comparison

The efficiency comparisons of ratio-type (d_{1R} and d_{2R}) and product-type (d_{1P} and d_{2P}) estimators have been made with $\bar{y}$, $\bar{y}_R$, t_{1R} and t_{2R}; and shown in Tables 4.1 and 4.2 respectively.

Table 4.1: *Comparison of mean squared errors of ratio-type estimators.*

Estimator	Mean squared error of	
	d_{1R}	d_{2R}
$\bar{y}$	if $K > \frac{\lambda_{(1)}}{2}$	if $K > \frac{\lambda_{(2)}}{2}$
$\bar{y}_R$	if $K < \frac{\left(1+\lambda_{(1)}\right)}{2}$	if $K < \frac{(1+\lambda_{(2)})}{2}$
t_{1R}	if $K > \frac{\left(1+\lambda_{(1)}C_1\right)}{2C_1}$	if either $K < \frac{(1+C_1\,\lambda_{(2)})}{2C_1}, \quad \lambda_{(2)} < \frac{1}{C_1}$ or $K > \frac{(1+C_1\,\lambda_{(2)})}{2C_1}, \quad \lambda_{(2)} > \frac{1}{C_1}$
t_{2R}	if $K > \frac{\left(1+\lambda_{(2)}C_2\right)}{2C_2}$	if either $K < \frac{(1+\,\lambda_{(2)}C_2)}{2C_2}, \quad \lambda_{(2)} < \frac{1}{C_2}$ or $K > \frac{(1+\lambda_{(2)}C_2)}{2C_2}, \quad \lambda_{(2)} > \frac{1}{C_2}$

Table 4.2: *Comparison of mean squared errors of product-type estimators.*

Estimator	Mean squared error of	
	d_{1P} is less than	d_{2P} is less than
$\bar{y}$	if $K < -\frac{\lambda_{(1)}}{2}$	if $K < -\frac{\lambda_{(2)}}{2}$
$\bar{y}_P$	if $K > -\frac{(1+\lambda_{(1)})}{2}$	if $K > -\frac{(1+\lambda_{(2)})}{2}$
t_{1P}	if $K < -\frac{(1+\lambda_{(1)}C_1)}{2C_1}$	if either $K < -\frac{1}{2}\frac{(1+\lambda_{(2)}C_1)}{C_1}$, $\lambda_{(2)} > \frac{1}{C_1}$ or $K > -\frac{1}{2}\frac{(1+\lambda_{(2)}C_1)}{C_1}$, $\lambda_{(2)} < \frac{1}{C_1}$
t_{2P}	if $K < -\frac{(1+\lambda_{(1)}C_2)}{2C_2}$	if either $K < -\frac{1}{2}\frac{(1+\lambda_{(2)}C_2)}{C_2}$, $\lambda_{(2)} > \frac{1}{C_2}$ or $K > -\frac{1}{2}\frac{(1+\lambda_{(2)}C_2)}{C_2}$, $\lambda_{(2)} < \frac{1}{C_2}$

Table 4.1 exhibits that the ratio type estimator d_{1R} is better than $\bar{y}$, $\bar{y}_R$, t_{1R} and t_{2R} if

$$\frac{\left(1+\lambda_{(1)}C_1\right)}{2C_1} \;<\; K \;<\; \frac{(1+\lambda_{(1)})}{2} \tag{4.1}$$

We also note that the estimator d_{1R} is more efficient than d_{2R} if

$$K > \frac{(\lambda_{(1)}+\lambda_{(2)})}{2} \tag{4.2}$$

It is observed from Table 4.1 that the product-type estimator d_{1P} is more efficient than $\bar{y}$, $\bar{y}_P$, t_{1P} and t_{2P} if

$$-\frac{(1+\lambda_{(1)})}{2} \;<\; K \;<\; -\frac{(1+\lambda_{(1)}C_1)}{2C_1} \tag{4.3}$$

Further it can be proved that the product-type estimator d_{1P} is better than the product-type estimator d_{2P} if

$$K < -\frac{(\lambda_{(1)}+\lambda_{(2)})}{2} \tag{4.4}$$

5. Unbiased versions of the suggested estimators

In this section we will obtain the unbiased versions of the suggested estimators in Section 2, using two well known procedures: (i) Interpenetrating subsamples design and (ii) Jack-knife technique.

5.1. Interpenetrating sub-sample design

Let the sample in the form of n independent interpenetrating subsamples be drawn. Let y_i and x_i be unbiased estimates of the population totals $Y(=N\overline{Y})$ and $X(=N\overline{X})$ respectively based on the i^{th} independent interpenetrating subsample, $i=1,2,\ldots,n$. We now consider following ratio and product-type estimators of the population mean $\overline{Y}$:

$$d_1=\overline{y}\left(\overline{A}/\overline{a}\right) \tag{5.1}$$

$$d_{1n}=\left(\overline{A}/n\right)\sum_{i=1}^{n}\left(y_i/a_i\right) \tag{5.2}$$

$$d_2=\overline{y}\left(\overline{B}/\overline{b}\right) \tag{5.3}$$

$$d_{2n}=\left(\overline{B}/n\right)\sum_{i=1}^{n}\left(y_i/b_i\right) \tag{5.4}$$

$$d_3=\overline{y}\left(\overline{a}/\overline{A}\right) \tag{5.5}$$

$$d_{3n}=\sum_{i=1}^{n}y_i\,a_i/\left(n\overline{A}\right) \tag{5.6}$$

$$d_4=\overline{y}\left(\overline{b}/\overline{B}\right) \tag{5.7}$$

and

$$d_{4n}=\sum_{i=1}^{n}y_i\,b_i/\left(n\overline{B}\right) \tag{5.8}$$

where $\overline{a}$, $\overline{b}$, $\overline{A}$, $\overline{B}$, a_i and b_i are same as defined in Section 2.

It is easy to verify that

$$B(d_{1n}) = nB(d_1) \tag{5.9}$$

$$B(d_{2n}) = nB(d_2) \tag{5.10}$$

$$B(d_{3n}) = nB(d_3) \tag{5.11}$$

and

$$B(d_{4n}) = nB(d_4) \tag{5.12}$$

Thus we get the following ratio and product-type unbiased estimators of $\overline{Y}$ as

$$d_{1u} = \frac{(nd_1 - d_{1n})}{(n-1)} \tag{5.13}$$

$$d_{2u} = \frac{(nd_2 - d_{2n})}{(n-1)} \tag{5.14}$$

$$d_{3u} = \frac{(nd_3 - d_{3n})}{(n-1)} \tag{5.15}$$

$$d_{4u} = \frac{(nd_4 - d_{4n})}{(n-1)} \tag{5.16}$$

The properties of these unbiased estimators (d_{ju}, $j = 1$ to 4) can be studied on the lines of Murthy and Nanjamma (1959).

Remark 5.1. In the case of simple random sampling without replacement (SRSWOR), let y_i and x_i denote respectively the y and x values of the sample of unit, $i = 1, 2, \ldots, n$. We have

$$d_1 = \bar{y}\left(\overline{A}/\bar{a}\right)$$

$$d_{1n} = \left(\overline{A}/n\right)\sum_{i=1}^{n}\left(y_i/a_i\right)$$

$$d_2 = \bar{y}\left(\overline{B}/\bar{b}\right)$$

$$d_{2n} = \left(\overline{B}/n\right) \sum_{i=1}^{n} \left(y_i/b_i\right)$$

$$d_3 = \bar{y}\left(\overline{a}/\overline{A}\right)$$

$$d_{3n} = \sum_{i=1}^{n} y_i a_i / \left(n\overline{A}\right)$$

$$d_4 = \bar{y}\left(\overline{b}/\overline{B}\right)$$

and

$$d_{4n} = \sum_{i=1}^{n} y_i b_i / \left(n\overline{B}\right)$$

It can be shown under SRSWOR scheme that the following ratio-type estimators are unbiased for population mean $\overline{Y}$ as

$$d_{1u}^{*} = \frac{n(N-1)}{N(n-1)}\bar{y}\left(\frac{\overline{A}}{\overline{a}}\right) - \frac{(N-n)}{N(n-1)}\frac{\overline{A}}{n}\sum_{i=1}^{n}\left(y_i/a_i\right) \tag{5.17}$$

$$d_{2u}^{*} = \frac{n(N-1)}{N(n-1)}\bar{y}\left(\frac{\overline{B}}{\overline{b}}\right) - \frac{(N-n)}{N(n-1)}\frac{\overline{B}}{n}\sum_{i=1}^{n}\left(y_i/b_i\right) \tag{5.18}$$

$$d_{3u}^{*} = \frac{n(N-1)}{N(n-1)}\bar{y}\left(\frac{\overline{a}}{\overline{A}}\right) - \frac{(N-n)}{N(n-1)}\frac{1}{n\overline{A}}\sum_{i=1}^{n} y_i a_i \tag{5.19}$$

$$d_{4u}^{*} = \frac{n(N-1)}{N(n-1)}\bar{y}\left(\frac{\overline{b}}{\overline{B}}\right) - \frac{(N-n)}{N(n-1)}\frac{1}{n\overline{B}}\sum_{i=1}^{n} y_i b_i \tag{5.20}$$

To the first degree of approximation, it can be shown that

$$Var\left(d_{1u}^{*}\right) = Var\left(d_{1R}\right) \tag{5.21}$$

$$Var\left(d_{2u}^{*}\right) = Var\left(d_{2R}\right) \tag{5.22}$$

$$Var\left(d_{3u}^{*}\right) = Var\left(d_{1p}\right) \tag{5.23}$$

and

$$Var\left(d_{4u}^{*}\right) = Var\left(d_{2p}\right). \tag{5.24}$$

Thus the unbiased estimators d_{1u}^{*}, d_{2u}^{*}, d_{3u}^{*} and d_{4u}^{*} are to be preferred over biased estimators d_{1R}, d_{2R}, d_{1p} and d_{2p} respectively.

5.2. Jack-knife technique

We may take $n = 2m$ and split the sample at random into two subsamples of m units each. Let $\bar{y}_i, \bar{x}_i$ $(i = 1,2)$ be unbiased estimators of population mean $\overline{Y}$ and $\overline{X}$ respectively based on the subsamples and $\bar{y}$, $\bar{x}$ the means based on the entire sample. Thus $\left(\bar{a}_i, \bar{b}_i; i = 1,2\right)$ are unbiased estimators based on the sub-samples and $\left(\bar{a}, \bar{b}\right)$ the means based on the entire sample i.e.,

$$\begin{aligned}
\bar{a}_i &= \left(x_M \bar{x}_i + x_m^2\right), \\
\bar{b}_i &= \left\{\left(x_M - x_m\right)\bar{x}_i + x_m^2\right\}, \\
\bar{a} &= \left(\bar{x} x_M + x_m^2\right),
\end{aligned}$$

and

$$\bar{b} = \left\{\left(x_M - x_m\right)\bar{x} + x_m^2\right\},$$

Thus motivated by Quenoulle (1956) we define the following ratio and product-type unbiased estimators of population mean $\overline{Y}$ as

$$d_{1J}^{(u)} = \frac{(2N-n)}{N} d_1 - \frac{(N-n)}{2N}\left\{d_1^{(1)} + d_1^{(2)}\right\} \tag{5.25}$$

$$d_{2J}^{(u)} = \frac{(2N-n)}{N} d_2 - \frac{(N-n)}{2N}\left\{d_2^{(1)} + d_2^{(2)}\right\} \tag{5.26}$$

$$d_{3J}^{(u)} = \frac{(2N-n)}{N} d_3 - \frac{(N-n)}{2N}\left\{d_3^{(1)} + d_3^{(2)}\right\} \tag{5.27}$$

and

$$d_{4J}^{(u)} = \frac{(2N-n)}{N} d_4 - \frac{(N-n)}{2N}\left\{d_4^{(1)} + d_4^{(2)}\right\} \tag{5.28}$$

where d_1, d_2, d_3 and d_4 are same as defined in Section 5, and

$$d_1^{(i)} = \bar{y}_i \left(\overline{A}/\overline{a}_i\right), \quad d_2^{(i)} = \bar{y}_i \left(\overline{B}/\overline{b}_i\right), \quad d_3^{(i)} = \bar{y}_i \left(\overline{a}_i/\overline{A}\right)$$

and

$$d_4^{(i)} = \bar{y}_i \left(\overline{b}_i/\overline{B}\right), \ (i = 1,2).$$

Following the procedure outlined in Sukhatme and Sukhatme [1970, pp. 161-165], it can be shown to the first degree of approximation that the variance expressions of $d_{lJ}^{(u)}$, $(l = 1,2,3,4)$ and variance expressions of d_{1R}, d_{2R}, d_{1p} and d_{2p} respectively are same.

Thus we advocate that one can prefer the unbiased estimators $d_{lJ}^{(u)}$, $(l = 1,2,3,4)$ as compared to biased estimators d_{1R}, d_{2R}, d_{1p} and d_{2p}.

6. Empirical study

6.1. When the variates y and x are positively correlated

To see the performances of the suggested estimators d_{1R} and d_{2R} over $\bar{y}$, $\bar{y}_R$, t_{1R} and t_{2R}, we have considered eight natural population data sets. Descriptions of the populations are given below:

Table 6.1: *Description of populations.*

Pop. No.	Source	N	n	Y	X	ρ	C_x	C_y	C_1	C_2	K
1	Sahoo and Swain (1987)	4	2	Unit: (0.2,0.6, 0.9,0.8)	Unit: (0.1,0.2, 0.3,0.4)	0.87	0.51	0.49	1.4	2.6	0.84
2	Murthy (1967), p. 422 (13-44)	12	4	Number of cattle (Survey)	Number of cattle (Census)	0.98	1.05	0.99	1.23	4.49	0.92
3	Murthy (1967), p. 398 (1-12)	12	4	Number of Absentees	Number of Workers	0.80	0.52	0.63	1.35	2.52	0.96
4	Panse and Sukhatme (1967), p. 118 (1-25)	25	10	Parental plot mean (mm)	Parental plant value (mm)	0.53	0.07	0.03	1.83	2.15	0.62
5	Panse and Sukhatme (1967), p. 118 (1-20)	20	8	Parental plot mean (mm)	Parental plant value (mm)	0.56	0.07	0.04	1.83	2.15	0.29
6	Panse and Sukhatme (1967), p. 118 (1-10)	10	4	Progeny mean (mm)	Parental plant value (mm)	0.44	0.07	0.05	1.92	2.13	0.31
7	Singh and Chaudhary p. 176 (1-10)	10	4	No. of Cows in milk (Survey)	No. of Cows in milk (Census)	0.97	0.63	0.58	1.26	2.81	0.89
8	Singh and Chaudhary p. 306	10	4	No. of inhabitants ('000) in 1980-81	No. of inhabitants ('000) in 1981-82	0.88	0.64	0.60	1.53	3.64	0.82
9	Samford (1962), p. 61 (1-9)	9	3	Acreage under oats in 1957	Acreage of crops and gross in 1947	0.07	0.10	0.29	1.86	2.12	0.19

To assess the biasedness of the ratio-type estimators $\bar{y}_R$, t_{1R}, t_{2R}, d_{1R} and d_{2R}, we have computed the following quantities for the population given in Table 6.1 using the formulae:

$$B_1 = \left| \frac{B(\bar{y}_R)}{\theta \bar{Y} C_x^2} \right| = |(1-K)| \tag{6.1}$$

$$B_2 = \left|\frac{B(t_{1R})}{\theta \overline{Y} C_x^2}\right| = \frac{1}{C_1}\left|\left(\frac{1}{C_1} - K\right)\right| \tag{6.2}$$

$$B_3 = \left|\frac{B(t_{2R})}{\theta \overline{Y} C_x^2}\right| = \frac{1}{C_2}\left|\left(\frac{1}{C_2} - K\right)\right| \tag{6.3}$$

$$B_4 = \left|\frac{B(d_{1R})}{\theta \overline{Y} C_x^2}\right| = \lambda_{(1)}\left|\left(\lambda_{(1)} - K\right)\right| \tag{6.4}$$

$$B_5 = \left|\frac{B(d_{2R})}{\theta \overline{Y} C_x^2}\right| = \lambda_{(2)}\left|\left(\lambda_{(2)} - K\right)\right| \tag{6.5}$$

The findings are listed in Table 6.2.

Table 6.2: *Values of B_1, B_2, B_3, B_4 and B_5.*

Values of B_i's $i = 1$ to 5	Population								
	1	2	3	4	5	6	7	8	9
B_1	0.1600	0.0826	0.0433	0.7399	0.7087	0.6951	0.1109	0.1767	0.8079
B_2	0.0898	0.0847	0.1602	0.1554	0.1397	0.1128	0.0781	0.1125	0.1852
B_3	0.1752	0.1547	0.2125	0.0946	0.0812	0.0772	0.1897	0.1507	0.1318
B_4	0.0628	0.0668	0.0178	0.2227	0.2091	0.1534	0.0708	0.0702	0.2460
B_5	0.0374	0.0657	0.0299	0.0175	0.0081	0.0209	0.0644	0.0489	0.0171

Table 6.2 exhibits that the proposed estimator d_{2R} has least bias for all data sets except in population III considered here. In population III, the proposed estimator d_{1R} has least bias. Using the following formulae:

$$PRE\,(\bar{y}_R, \bar{y}) = \frac{MSE\,(\bar{y})}{MSE\,(\bar{y}_R)} \times 100 = \left[1 + \left(\frac{C_x}{C_y}\right)^2 (1 - 2K)\right]^{-1} \times 100 \tag{6.6}$$

$$PRE\,(t_{1R}, \bar{y}) = \frac{MSE\,(\bar{y})}{MSE\,(t_{1R})} \times 100 = \left[1 + \frac{1}{C_1}\left(\frac{C_x}{C_y}\right)^2 \left(\frac{1}{C_1} - 2K\right)\right]^{-1} \times 100 \tag{6.7}$$

$$PRE\,(t_{2R}, \bar{y}) = \frac{MSE\,(\bar{y})}{MSE\,(t_{2R})} \times 100 = \left[1 + \frac{1}{C_2}\left(\frac{C_x}{C_y}\right)^2 \left(\frac{1}{C_2} - 2K\right)\right]^{-1} \times 100 \tag{6.8}$$

$$PRE\,(d_{1R},\bar{y}) = \frac{MSE(\bar{y})}{MSE(d_{1R})} \times 100 = \left[1 + \left(\frac{C_x}{C_y}\right)^2 \lambda_{(1)}(\lambda_{(1)} - 2K)\right]^{-1} \times 100 \quad (6.9)$$

and

$$PRE\,(d_{2R},\bar{y}) = \frac{MSE(\bar{y})}{MSE(d_{2R})} \times 100 = \left[1 + \left(\frac{C_x}{C_y}\right)^2 \lambda_{21)}(\lambda_{(2)} - 2K)\right]^{-1} \times 100 \quad (6.10)$$

We have computed the percent relative efficiencies (PREs) of $\bar{y}_R$, t_{1R}, t_{2R}, d_{1R} and d_{2R} with respect to usual unbiased estimator $\bar{y}$ and compiled in Table 6.3.

Table 6.3: *Percent relative efficiencies of $\bar{y}_R$, t_{1R}, t_{2R}, d_{1R} and d_{2R} with respect to $\bar{y}$.*

Estimator	$PRE(.,\bar{y})$								
	Population								
	1	2	3	4	5	6	7	8	9
$\bar{y}$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\bar{y}_R$	383.33	2279.92	273.92	33.62	39.24	55.15	1263.21	380.08	92.90
t_{1R}	399.65	2063.93	252.32	94.69	107.82	110.63	1313.15	382.20	98.99
t_{2R}	218.13	169.80	161.47	112.07	125.30	115.91	249.18	175.31	99.49
d_{1R}	419.76	2421.29	274.71	78.95	90.93	104.63	1408.39	426.60	98.41
d_{2R}	425.54	2430.62	274.35	136.19	145.40	120.65	1428.98	432.85	100.41

Table 6.3 shows that the proposed estimator d_{2R} has largest gain in efficiency for all population data sets except in population III, where the proposed estimator d_{1R} has maximum gain in efficiency. We also note that the proposed estimator d_{1R} dominates over the estimators ($\bar{y}$, $\bar{y}_R$, t_{1R} and t_{2R}) in population I, II, III, IV, VII and VIII. Thus the proposed estimators d_{1R} and d_{2R} are to be preferred over other estimators.

Finally, from Tables 6.2 and 6.3 we recommend the use of the proposed estimator d_{2R} in practice as it has largest gain in efficiency and also fewer bias in all population data sets except in population III, where the proposed estimator d_{1R} has largest gain in efficiency as well as less bias and hence d_{1R} is to be recommended for this population data set.

6.2. When the variates y and x are negatively correlated

To assess the biasdeness and efficiency of the product-type estimators $\bar{y}_p$, t_{1p}, t_{2p}, d_{1p} and d_{2p} we have considered natural population data sets.

Table 6.4: *Description of the populations.*

Pop. No.	Source	N	n	Y	X	ρ	$\mathbf{C}_x$	$\mathbf{C}_y$	C_1	C_2	K
1	Maddla, G.S. (1977), p. 96	16	4	Capita Consumption	Deflated price	−0.97	0.24	0.17	1.68	2.39	−0.68
2	Gupta, S.P. and Gupta, A. (1999) p. 65	5	2	Artificial Population		−0.96	0.52	0.51	1.43	2.74	−0.93

To observe the biasedness of the estimators $\bar{y}_p$, t_{1p}, t_{2p}, d_{1p} and d_{2p}, we use the following formulae:

$$B_1^* = \left|\frac{B(\bar{y}_p)}{\theta \overline{Y} C_x^2}\right| = |K| \tag{6.11}$$

$$B_2^* = \left|\frac{B(\bar{y}_{1p})}{\theta \overline{Y} C_x^2}\right| = \left|\frac{K}{C_1}\right| \tag{6.12}$$

$$B_3^* = \left|\frac{B(t_{2p})}{\theta \overline{Y} C_x^2}\right| = \left|\frac{K}{C_2}\right| \tag{6.13}$$

$$B_4^* = \left|\frac{B(d_{1p})}{\theta \overline{Y} C_x^2}\right| = \lambda_{(1)}|K| \tag{6.14}$$

$$B_5^* = \left|\frac{B(d_{2p})}{\theta \overline{Y} C_x^2}\right| = \lambda_{(2)}|K| \tag{6.15}$$

The quantities $B*'s\,(i = 1 \text{ to } 5)$ have been computed and findings are given in Table 6.5.

Table 6.5: *Values of B_1^*, B_2^*, B_3^*, B_4^* and B_5^*.*

Population	Values of B_i^*'s, i = 1 to 5				
	B_1^*	B_2^*	B_3^*	B_4^*	B_5^*
1	0.6814	0.4043	0.2843	0.5099	0.4104
2	0.9338	0.6508	0.3409	0.8422	0.8156

Using the following formulae:

$$PRE\,(\bar{y}_p,\bar{y}) = \frac{MSE(\bar{y})}{MSE(\bar{y}_P)} \times 100 = \left[1+\left(\frac{C_x}{C_y}\right)^2 (1+2K)\right]^{-1} \times 100 \qquad (6.16)$$

$$PRE\,(t_{1p},\bar{y}) = \frac{MSE(\bar{y})}{MSE(t_{1P})} \times 100 = \left[1+\left(\frac{C_x}{C_y}\right)^2 \frac{1}{C_1}\left(\frac{1}{C_1}+2K\right)\right]^{-1} \times 100 \qquad (6.17)$$

$$PRE\,(t_{2p},\bar{y}) = \frac{MSE(\bar{y})}{MSE(t_{2P})} \times 100 = \left[1+\left(\frac{C_x}{C_y}\right)^2 \frac{1}{C_2}\left(\frac{1}{C_2}+2K\right)\right]^{-1} \times 100 \qquad (6.18)$$

$$PRE\,(d_{1p},\bar{y}) = \frac{MSE(\bar{y})}{MSE(\bar{y}_P)} \times 100 = \left[1+\left(\frac{C_x}{C_y}\right)^2 \lambda_{(1)}\left(\lambda_{(1)}+2K\right)\right]^{-1} \times 100 \qquad (6.19)$$

and

$$PRE\,(d_{2p},\bar{y}) = \frac{MSE(\bar{y})}{MSE(\bar{y}_P)} \times 100 = \left[1+\left(\frac{C_x}{C_y}\right)^2 \lambda_{(2)}\left(\lambda_{(2)}+2K\right)\right]^{-1} \times 100 \qquad (6.20)$$

We have computed the percent relative efficiencies (PREs) of $\bar{y}_p$, t_{1p}, t_{2p}, d_{1p} and d_{2p} with respect to usual unbiased estimator $\bar{y}$ and the results are shown in Table 6.6.

Table 6.6: *Percent relative efficiencies of $\bar{y}_p$, t_{1p}, t_{2p}, d_{1p} and d_{2p} with respect to $\bar{y}$.*

Estimators		$\bar{y}$	$\bar{y}_p$	t_{1p}	t_{2p}	d_{1p}	d_{2p}
$PRE(.,\bar{y})$	Population 1	100.00	390.97	1578.36	524.73	1764.62	1658.49
	Population 2	100.00	1133.69	701.62	236.13	1181.21	1143.86

Tables 6.5 and 6.6 show that the proposed estimators d_{1p} and d_{2p} are more efficient (with substantial gain) than usual unbiased estimator $\bar{y}$, product estimator $\bar{y}_p$ and the estimators t_{1p} and t_{2p} reported by Sahoo and Mohanty (1995), but these two estimators (d_{1p} and d_{2p}) are more biased than t_{1p} and t_{2p}. Thus if the variance / MSE's criterion of judging the performance of the estimators are adopted and also the biasedness of the estimators are not of primary concern then the proposed estimators d_{1p} and d_{2p} are recommended for their use in practice.

Acknowledgement

Authors are thankful to the referee for his valuable suggests ions regarding improvement of the earlier draft of the paper.

References

Bandyopadhyaya, S. (1980). Improved ratio and product estimators. *Sankhya*, 42,C, 45-49.

Cochran, W. G. (1963). *Sampling Techniques*. John Wiley and Sons Inc., New York, II Edition.

Gupta, P. C. (1978). On some quadratic and higher degree ratio and product estimators. *Journal of the Indian Society of Agricultural Statistics*, 30, 71-80.

Gupta, S. P. and Gupta, A. (1999). *Statistical Methods*. Sultan Chand & Sons, 6.5, New Delhi.

Maddala, G. S. (1977). *Econometrics*. McGraw Hills pub.Co., New York.

Mohanty, S. and Das, M. N. (1971). Use of transformation in sampling. *Journal of the Indian Society of Agricultural Statistics*, 23, 2, 83-87.

Mohanty, S. and Sahoo, J. (1995). A note on improving the ratio method of estimation through linear transformation using certain known population parameters. *Sankhya*, 57, B, 93-102.

Murthy, M. N. (1964). Product method of estimation. *Sankhya*, 69-74.

Murthy, M. N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta.

Murthy, M. N. and Nanjamma, N. S. (1959). Almost unbiased ratio estimators based on interpenetrating sub samples. *Sankhya*, 21, 381-392.

Naik, V. D. and Gupta, P. C. (1991). A general class of estimators for estimating population means using auxiliary information. *Metrika*, 38, 11-17.

Panse, V. G. and Sukhatme, P. V. (1967). Statistical methods for agricultural workers. *Indian Council of Agricultural Research*, New Delhi.

Reddy, V. N. (1973). On ratio and product methods of estimation. *Sankhya*, B, 35, 307-317.

Reddy, V. N. (1974). On a transformed ratio method of estimation. *Sankhya*, C, 36, 59-70.

Robson, D. S. (1957). Application of multivariate Polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 59, 1225-1226.

Rueda, M. and González, S. (2008). A new ratio-type imputation with random disturbance. *Applied Mathematics Letters*, 21, 9, 978-982.

Sahai, A. and Sahai, A. (1985). On efficient use of auxiliary information. *Journal of Statistical Planning and Inference*, 12, 203-212.

Sahoo, L. N. and Swain, A. K. P. C. (1987). Some modified ratio estimators. *Meteron*, 286-292.

Sampford, M. R. (1962). *An Introduction to Sampling Theory*. Oliver and Boyd.

Samawi, H. M. and Al-Saleh, M. F. (2007). On bivariate ranked set sampling for ratio and regression estimators. *International Journal of Modelling and Simulation*, 27, 4, 299-305.

Singh, D. and Chaudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*. Wilely eastern Lt., New Delhi.

Singh, S. (2003). *Advanced Sampling Theory with Applications. How Michael "selected" Amy, 1-1247*, Kluwer Academic Publishers, The Netherlands.

Sisodia, B. V. and Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics*, 33, 13-18.

Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys. *Cal. Statist. Assoc. Bull.*, 6, 121-132.

Srivastava, S. K. (1983). Predictive estimation of finite population mean using product estimator. *Metrika*, 30, 93-99.

Srivenkataramana, T. (1978). Change of origin and scale in ratio and difference method of estimation in sampling. *The Canadian Journal of Statistics*, 6,1, 79-86.

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika*, 67, 194-204.

Sukhatme, P. V. and Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications*, 2nd ed. Ames. Iowa State University Press.

Upadhyaya, L. N. and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* 41, 5, 627-636.

Vos, J. W. E. (1980). Mixing of direct ratio and product method estimators. *Statistica Neerlandica*, 34, 209-213.

Walsh, J. E. (1970). Generalization of ratio estimate for population total. *Sankhya*, A., 32, 99-106.

SORT 34 (2) July-December 2010, 181-200

On the use of simulation methods to compute probabilities: application to the first division Spanish soccer league

Ignacio Díaz-Emparanza* and Vicente Núñez-Antón**

Universidad del País Vasco

Abstract

We consider the problem of using the points a given team has in the First Division Spanish Soccer League to estimate its probabilities of achieving a specific objective, such as, for example, staying in the first division or playing the European Champions League. We started thinking about this specific problem and how to approach it after reading that some soccer coaches indicate that a team in the first division guarantees its staying in that division if it has a total of 42 points at the end of the regular season. This problem differs from the typical probability estimation problem because we only know the actual cumulative score a given team has at some point during the regular season. Under this setting a series of different assumptions can be made to predict the probability of interest at the end of the season. We describe the specific theoretical probability model using the multinomial distribution and, then, introduce two approximations to compute the probability of interest, as well as the exact method. The different proposed methods are then evaluated and also applied to the example that motivated them. One interesting result is that the predicted probabilities can then be dynamically evaluated by using data from the current soccer competition.

MSC: 62F05, 62P99, 6204, 6207

Keywords: Monte Carlo simulations, multinomial distribution, prediction, soccer league

Address for correspondence: Vicente Núñez-Antón, Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco, Avenida Lehendakari Aguirre, 83, E-48015 Bilbao, Spain. Phone: +34 94 601 3749; Fax: +34 94 601 3754; E-mail: vicente.nunezanton@ehu.es

* Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco, Bilbao, Spain. E-mail: ignacio.diaz-emparanza@ehu.es

** Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco, Bilbao, Spain. E-mail: vicente.nunezanton@ehu.es

Received: January 2010

Accepted: July 2010

1. Introduction

The *Liga de Fútbol Profesional* (LFP) stated that, starting at the regular season 1997-1998, there will be twenty teams competing in the First Division Spanish Soccer League. During the regular season, each team should play two games against each one of the remaining nineteen teams, one game at its own field and the other one at the other team's field. Therefore, during the regular season there will be a total of thirty-eight games played by each one of the teams participating in this league. After the 1995-1996 regular season the LFP stated the actual scoring system: a win gets a team three points, a draw one point and a defeat, no points. In this way, at the end of the regular season (i.e., after the thirty-eight games have been played), teams classified in the last three positions in the table (i.e., positions eighteenth to twentieth) will lose their place in the first division and will have to play the next regular season in the Second Division Spanish Soccer League. In addition, teams classified in the first four positions will play the European Champions League, the most prestigious soccer tournament in Europe (i.e., the one that only "the best" soccer teams in Europe will play), while teams classified in the fifth and sixth positions will play the UEFA tournament (nowadays called *Europa League*), an important soccer tournament for the so called "next-to-the-best" teams in Europe.

Soccer is the most important sport in Spain and there are several sports-related (TV and radio) programs that concentrate most of their attention and efforts on the Spanish soccer league. It is a fact that sports-related programs in Spain can very well be labelled as "soccer-centred programs." In the past few years, it has been very frequent to hear sport broadcasters indicating that "soccer coaches training teams in the first division believe that if a team obtains a total of 42 points at the end of the regular season, the team will remain in this division for the next regular season."[1] That is, the 42 points figure somehow represents the barrier that will determine if a team plays in the first or second division for the next regular season. In fact, after reading a specialized and well known sports newspaper, the first author's older son asked him the question of why some specific soccer coaches indicate that a team in the first division guarantees its staying in that division if it has a total of 42 points at the end of the regular season. This very simple and straightforward question originated our curiosity as to how one can propose some kind of approach to answer it. It also made us ask ourselves whether the question had been raised before by someone else.

A very simple analysis of the available data for the last twelve regular seasons reveals that only in three occasions a team obtaining a total of 42 points at the end of the regular season lost its right to play in the first division for the next season. More specifically, during the 1999-2000 and the 2008-2009 regular seasons, Betis, a soccer team from Seville and, during the 2007-2008 regular season, Zaragoza, another soccer team from a city with the same name, were both sent to play the corresponding next regular season in

1. For more details on this see, for example: http://bit.ly/biPsGx, http://bit.ly/maportugal, http://bit.ly/brindis-osasuna, or http://bit.ly/racing42.

the second division. Even though there are some statistical papers that propose to model scores in soccer leagues (see, e.g., Lee, 1997; Karlis and Ntzoufras, 2000; Rue and Salvensen, 2000; Brillinger, 2008; and Karlis and Ntzoufras, 2009), we are not aware of any scientific study or attempt that has tried to find out the actual probability that a team playing in the first division with 42 points at the end of the regular season stays in that division during the next season. Along the same lines, we do not know of any attempt that has been able to establish, using a rigorous statistical reasoning or tool, the total number of points a team should obtain during the regular season, so that it can stay in the first division for the next season. These represent two of the main objectives that have led us to put forward some of the proposals included in the next sections.

The rest of the paper is organized as follows. Section 2 introduces some basic notation and contains a brief description of all of the possible classifications at the end of a regular season for a four and a twenty team league. Section 3 describes the use of the multinomial distribution in the context of the soccer league under study, as well as the normal approximation, Monte Carlo simulations approximation and the exact probability computation for the different probabilities of interest. In Section 4, we include the proposed method to compute the probability of a team staying in the first division for the next regular season. Section 5 puts forward a dynamic probability computation method that allows the researcher or individual to compute different probabilities of interest during the regular season. Finally, Section 6 ends with some conclusions and practical recommendations. All of the proposals contained in the different sections of the paper are illustrated and evaluated with data from the First Division Spanish Soccer League.

2. Basic notation and possible classifications settings

Let A_i represent each of the i ($i = 1,\ldots,N$) teams participating in a given league. That is, all teams will be denoted by $A_1,\ldots,A_N$. The order of the team is not relevant and it could be, in fact, alphabetical, per region, or sorted by any other criteria. Let E_{ik} represent the points obtained by team i on its k-th game during the regular season ($k = 1,\ldots,2(N-1)$). In this way, the result of the game played by teams A_i and A_j in the k-th of the regular season can be easily summarized by the 2×1 score vector $(E_{ik},E_{jk})'$. In the following sections we will analyze these results for the case of a league of four and twenty teams, which is the actual size of the First Division Spanish Soccer League under study.

As we will see in later sections and without loss of generality, we assume equiprobability. That is, in each game we assume that the probability that the local team wins, loses or that the result is a draw are all equal. This implies that all possible final classifications have the same probability. This assumption implies that no additional a priori information is needed to be able to compute, for example, the probability that a team loses its category when having 42 points at the end of the regular season, or the probabil-

ity of winning the league with a given number of points at the end of the regular season. In this sense, all of the results reported here could be applied not only to the First Division Spanish Soccer League, but also to any second division or to any other division or league using the scoring system proposed in this league. In any case and given that it is very unlikely that all teams in this league have the same constant probability of winning a given soccer game, this is clearly a restrictive hypothesis that may be considered too strong in some cases for practical reasons but, at the same time, it may also be considered simple enough to be interesting from a didactic point of view. In fact, this is the main reason to start analyzing this problem under this assumption because, in our view, it clearly simplifies its solution and, in addition, it will also provide reference values for the probabilities of interest that may then be useful for the analysis of any other soccer league one wishes to study in the future.

A less restrictive assumption that also allows us to obtain interesting statistical results, can be that of *equal strength*. In order for this assumption to hold, the probability that the local team wins and the probability that it loses should be the same. That is, if we let p_1 be the probability that the local team wins, p_2 the probability that the result is a draw, and p_3 the probability that the local team loses the game, *equal strength* will occur if $p_1 = p_3 = (1-p_2)/2$. One interesting fact about this assumption is that it includes the equiprobability case as a particular case (i.e., if $p_1 = p_2 = p_3 = 1/3$), but it also includes additional possibilities that could also be analyzed. Along these lines, if we consider the First Division Spanish Soccer League historical data for the 11,242 games played from the 1976-1977 up to the 2008-2009 seasons, the estimated value we obtain for p_2, if we use the relative frequency for the event that the result of the game is a draw is, approximately, $\hat{p}_2 = 0.25$. In the following sections, we will use both the equiprobability and the *equal strength* assumptions. Finally, we should also mention that the equiprobability and equal strength assumptions imply that the probability that a given team wins, loses or that the result of its game is a draw, does not depend on which team it is playing against and that, therefore, there is an underlying independence assumption between games. This may also be a restrictive assumption but, in our view, it simplifies the solution to the problem of interest and, in addition, it provides the reader some very useful insights about the solution to a more complex problem.

2.1. A four-team soccer league

If we have four teams in the league, A_1, A_2, A_3, A_4, there will be three games in which a given team plays at home and three games in which it plays away from it, as a visiting team. That is, the regular season will have a total of six games. In this case, each date for which games are scheduled will have two games being played at the same time. If we let $a = (E_{11}, E_{21})'$ be the score for the game played by teams A_1 and A_2, and $b = (E_{31}, E_{41})'$ be the game played by teams A_3 and A_4, the possible scores for the first set of games to be played is listed in Table 1. After this first set of games is played, there are $3^2 = 9$ possible score vectors that are listed in the corresponding columns of Table 2. In order

Table 1: *Possible score vectors for a two-team soccer league.*

Possibilities	1	2	3
Score vector a:	$(3,0)'$	$(1,1)'$	$(0,3)'$
Score vector b:	$(3,0)'$	$(1,1)'$	$(0,3)'$

Table 2: *Possible scores for a four-team soccer league after the first set of games have been played.*

Result	a1b1	a1b2	a1b3	a2b1	a2b2	a2b3	a3b1	a3b2	a3b3
Team A_1	3	3	3	1	1	1	0	0	0
Team A_2	0	0	0	1	1	1	3	3	3
Team A_3	3	1	0	3	1	0	3	1	0
Team A_4	0	1	3	0	1	3	0	1	3

to better understand both the notation and contents in Table 2, let us describe one of the results provided therein (i.e., axby). The result in the fourth column of Table 2 (i.e., a2b1) indicates that in the game between teams A_1 and A_2 the result was a draw (i.e., the second possible result for the score vector a in Table 1, or a2), and in the game between teams A_3 and A_4 the result was that team A_3 won (i.e., the first possible result for the score vector b in Table 1, or b1). For this specific case, the final scores obtained by each of the teams A_1, A_2, A_3 and A_4 after the two games have been played would be of 1, 1, 3, and 0 points, respectively and, thus, the score vector would then be $(1,1,3,0)'$ (see the fourth column in Table 2). In summary, after the regular season ends (i.e., after each team has played its six corresponding games), there could be $3^{2\times 6} = 9^6 = 531,441$ *different*[2] results that will, in turn, generate their corresponding score vectors for these four teams.

The aforementioned number of possible results is clearly quite large. However, it can be easily managed by a computer. As the reader may have already guessed, this is exactly the situation in the first round of the Champions League competition, where only the first two teams in each group of four teams advance to the next round. Therefore, it would not be difficult to compute, for example, what would be the exact probability, for each possible score, that a team finishes the competition in the first two positions (i.e., the probability that the team advances to the next round in the Champions League):

Score:	≤ 6	7	8	9	≥ 10
Prob(next round):	0	0.0047477	0.18593	0.97050	1

That is, in all of the possible $531,441$ *different* results, we look for all results where a team having a given score finishes the competition in the first two positions (i.e., these

2. These results are different in the sense that, even though the scores could end up being equal for some of the cases, they were generated from different results in the games the teams have played.

will the favourable cases) and divide this absolute frequency by the total number of cases where teams had obtained this score (i.e., these will be the possible cases).

2.2. The twenty-team or first division Spanish soccer league

The Spanish First Division Soccer League, as well as, for example, the ones in France or the United Kingdom, has a total of twenty teams (i.e., $N = 20$), so that every round there will be ten different games played at the same time. In addition, every team should play nineteen games at home and another nineteen games as a visiting team, so that the regular season will have a total of thirty-eight different rounds.

If we let $(E_{ik}, E_{jk})'$ be the score vector representing the result of the game between teams A_i and A_j in the k-th round of games during the regular season, we have that:

$$(E_{ik}, E_{jk}) = \begin{cases} (3,0) & \text{if } A_i \text{ wins} \\ (1,1) & \text{if the result is a draw} \\ (0,3) & \text{if } A_j \text{ wins} \end{cases}$$

Therefore, after the k-th round of games is over (i.e., after the ten scheduled games have been played by the twenty teams in the league), we will have that $E_k = (E_{1k}, E_{2k}, \ldots, E_{20k})'$ represents the score vector assigned to all teams for the games played that date. Moreover and given that for each one of the ten games played that date there are only three possible different results, the number of different score vectors that one can obtain for that specific date is equal to $3^{10} = 59,049$.

Let C_k be the 20×1 score vector containing the sum of the scores from the first up to the k-th round of games, so that

$$C_k = \sum_{l=1}^{k} E_l,$$

and $C_k = (C_{1k}, \ldots, C_{20k})'$. Therefore, $C_{ik},\ i = 1, \ldots, 20;\ k = 1, \ldots, 38$ represents the score team A_i has after playing k games. If we place the elements of C_k in descending order and denote this new score vector by $C_k^o = (C_{(1)k}, \ldots, C_{(20)k})' = (C_{1k}^o, \ldots, C_{20k}^o)'$, we will have in the elements of the ordered score vector C_k^o the complete information about *teams classification or standings after k games have been played*, which will be very relevant to compute the probabilities of interest.

If, for example, we wish to analyze the number of vectors with possible *different* scores after two rounds of games have been played (i.e., C_2), we have to consider that each one of the 59,049 resulting score vectors for the second date for which games were scheduled can be added to each one of the 59,049 score vectors for the first date for which games were also scheduled, making a total of $3^{10\times 2} = 59,049 \times 59,049 = 3,486,784,401$ possible results, even though we know that a large number of them will be basically equal. If we follow the same reasoning, we can find out that the number of

vectors with possible *different* scores for the score vector C_{38} at the end of the regular season would then be $3^{10\times 38} = 3^{380} \simeq 2.023376E+181$.

In order to be able to compute the exact probability of losing the category for a team having 42 points, just as we did in Section 2.1 for the probability of advancing to the next round in the Champions League for the four teams' case, we would have to find out for how many of these $2.023376E+181$ score vectors a team having 42 points stays away from the last three positions in the table (i.e., stays away from the last three positions in the ordered score vector C^o_{38}). It is clear that, even with the current capabilities large computers have to compute this probability, it is not reasonable to think about working with such a large number of possibilities. For example, if the computer is able to compute 1,000 score vectors per second, after a year of computations, the computer would have only computed about $3.1536E+10$ score vectors. Therefore, there is a need to look for efficient and reasonable proposals that can make such a complicated computation of probabilities possible.

3. Multinomial distribution

The settings we have introduced in the previous sections allow us to state that, for each game and team, the set of possible results can be classified in the disjoint events: R_1 (winning the game), R_2 (game ends in a draw) or R_3 (losing the game). We now define the probabilities for these events as follows:

$$\Pr(R_j) = p_j \qquad \text{with} \quad 0 < p_j < 1, \quad j = 1,2,3 \quad \text{and} \quad p_1 + p_2 + p_3 = 1$$

To start with a simple setting, we can consider a discrete uniform probability distribution for the three alternatives, so that it is assumed that $p_1 = p_2 = p_3 = 1/3$. That is, we start with the initial aforementioned equiprobability assumption. Under this assumption, for any team in the league, the random variable $X = (X_1, X_2, X_3)'$, describing the event that after n dates for which games were scheduled during the regular season, there were x_1 times where the event R_1 occurred, x_2 times where the event R_2, and x_3 times where the event R_3 occurred, follows a multinomial distribution with probability mass function given by (see, e.g., Morris, 1975)

$$\Pr(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!}\, p_1^{x_1} p_2^{x_2} p_3^{x_3}, \tag{1}$$

$$x_1 + x_2 + x_3 = n, \quad 0 < p_j < 1, \quad j = 1,2,3 \quad \text{and} \quad p_1 + p_2 + p_3 = 1$$

It is well known that the marginal distribution of each of the X_j variables from a multinomial distribution follows a binomial distribution with parameters n and p_j; that is, $X_j \sim B(n, p_j)$, $j = 1,2,3$ with $E(X_j) = np_j$, $Var(X_j) = np_j(1-p_j)$ and $Cov(X_i, X_j) =$

$-np_ip_j,\ i,j=1,2,3,\ i\neq j$. As we have already mentioned, we are under the equiprobability assumption. However, the distribution of the random variable $X=(X_1,X_2,X_3)'$ is multinomial as long as the assumed probabilities p_1, p_2 and p_3 remain unchanged for all games in the league. This implies that, for example, under the *equal strength* assumption, the distribution of the random variable X is also multinomial.

3.1. Normal distribution approximation

If we use the multivariate normal central limit (see, e.g., Agresti, 1990, p. 424; or Rao, 1973, p. 128), we can see that the multinomial distribution converges to the multivariate normal distribution, so that $X=(X_1,X_2,X_3)'$ converges in distribution to a multivariate normal distribution with mean vector given by $\mu_X=(np_1,np_2,np_3)'$ and variance-covariance matrix Σ_X with elements given by:

$$\Sigma_{X_{ij}}=\begin{cases} np_i(1-p_i) & \text{if} \quad i=j \\ -np_ip_j & \text{if} \quad i\neq j \end{cases} \tag{2}$$

The score a given team A_i obtains after playing n games is $C_{in}=3X_1+X_2$, a linear combination of the components of the asymptotic multivariate normal random variable X, which can be written as $C_{in}=d'X$, with $d'=(3,1,0)$. Therefore, C_{in} converges to the univariate normal distribution $N(d'\mu_X,d'\Sigma_Xd)$.

For the specific case under study, we have that $n=38$ and $p_1=p_2=p_3=1/3$, so that the standard conditions (i.e., $np_i>5$ and $n(1-p_i)>5$, see, e.g., Hogg and Tanis, 1988 or Cryer and Miller, 1991) for a valid approximation hold and, therefore, we have that

$$C_{i38}\approx N(50.67,59.11) \tag{3}$$

The probability that a given team in the league loses its category is the probability that its ordered position after the regular season ends in one of the last three out of the twenty possible positions. Thus, we are interested in computing the critical score value, say C_{38c}, such that there would be three teams below it (around 15% or 3 out of 20 teams) or three teams having a score smaller than C_{38c}. In other words, C_{38c} should be such that $\Pr(C_{i38}\leq C_{38c})\geq 0.15$. In order to compute C_{38c} and using the result in (3), we have that

$$\frac{C_{i38}-50.67}{\sqrt{59.11}}\approx N(0,1)$$

Therefore,

$$\Pr\left(\frac{C_{i38}-50.67}{\sqrt{59.11}}\leq -z_{0.15}=-1.0364\right)=0.15 \tag{4}$$

being $-z_{0.15} = -1.0364$, the 15-th percentile of the standard normal distribution, $N(0,1)$. Solving for C_{i38} in the left hand side inside the parenthesis, leads us to obtain that

$$\Pr\left[C_{i38} \leq 50.67 - (1.0364)\sqrt{59.11}\right] = 0.15$$

and, thus, $\Pr(C_{i38} \leq 42.70) = 0.15$.

If we apply a standard continuity correction[3], we would have that the C_{38c} value we are searching for is $C_{38c} = 43$. This value leads us to obtain that $\Pr(C_{i38} \leq 43) = 0.1755$. Moreover, we can also easily verify that $\Pr(C_{i38} \leq 42) = 0.1439$. Therefore, the objective score for any team wishing not to lose its category in the First Division Spanish Soccer League should be of 43 points.

In addition, if we use the *equal strength* assumption with a probability that the result of a draw is $p_2 = 0.25$, we have that $C_{i38} \approx N(52.25, 65.92)$ and, therefore, $\Pr(C_{i38} \leq 43.83) = 0.15$. If we use again the aforementioned continuity correction and given that we can easily compute $\Pr(C_{i38} \leq 44) = 0.1699$ and $\Pr(C_{i38} \leq 43) = 0.1406$, we would now have that $C_{38c} = 44$.

3.2. Monte Carlo simulations approximation

A second alternative approach to obtain the distribution of C_{i38} consists of using a simulations approach. Let us begin by recalling that we are assuming equal probabilities for each one of the three possible results than a given game can have; that is, R_1 (winning the game), R_2 (game ends in a draw) and R_3 (losing the game):

$$\Pr(R_j) = p_j = \frac{1}{3} \qquad \text{for} \quad j = 1,2,3.$$

Most statistical packages include random number generators based on the uniform distribution and, thus, it is quite simple to simulate the result of a given game with the use of this software[4]. In this sense, if we assume independence among the games played at each round during the regular season, the results for ten independent games can be easily simulated in order to obtain the scores for all twenty teams after that specific date. We also assume that the probabilities p_j remain constant for each game so that, under the previous assumptions, the results for the different round of games are also independent. We can then repeat the whole simulation process thirty-eight times in order to be able to simulate the results for the final standings for all twenty teams in the league at the end of the regular season. The whole process can be easily summarized as follows:

---

3. We consider that for any $x \in \{0,1,\ldots,n\}$, if the conditions to consider the normal approximation a valid one hold, then $\Pr(X \leq x) = \Pr(X < x+1)$ can be well approximated by $\Pr(Y \leq x + \frac{1}{2})$, where Y is a normal random variable having the same mean and variance as the random variable X.

4. We have used the open source software package `gretl` (see, e.g., `http://gretl.sourceforge.net` or Cottrell and Lucchetti, 2009).

1. Based on the uniform distribution, we generate the results of the game between teams A_i and A_j in the first date for which games are scheduled, and obtain the corresponding score vector $(E_{i1}, E_{j1})'$.
2. Repeat step 1 ten times and obtain the results for all ten games played in the first date for which games are scheduled. At the end of this step, we obtain the 20×1 score vector $E_1 = (E_{11}, \ldots, E_{20,1})'$.
3. Repeat steps 1 and 2 for each one of the thirty-eight dates for which games are scheduled, generating the corresponding 20×1 score vectors $E_k = (E_{1k}, \ldots, E_{20k})'$, $k = 1, \ldots, 38$, and obtain the sum of the scores for all twenty teams after the "simulated" regular season ends; that is, obtain the 20×1 final scores vector $C_{38} = \sum_{l=1}^{38} E_l$.
4. Finally, repeat steps 1 to 3 a large number of times, say M, and obtain the simulated frequency distribution for $C_{38} = (C_{1,38}, \ldots, C_{20,38})'$, an approximation of the probability distribution for this random variable and, accordingly, of its individual components C_{i38}.

If we follow the procedure described in Díaz-Emparanza (2002, equation (8)), we see that, with a 95% confidence level, M =10,000 replications will suffice to guarantee a precision of ± 0.007 in the estimation of the 15% distribution percentile of interest.

After these simulations are performed, we can straightforwardly obtain that $\Pr(C_{i38} \leq 43) = 0.1770$ and $\Pr(C_{i38} \leq 42) = 0.1448$, values that, as can be easily verified, are very close to those obtained in Section 3.1 with the use of the normal approximation.

However, it is also possible to consider an alternative interpretation of the results obtained in this simulation approach. In Section 2.2 we have indicated that there is a large number of *different* possibilities for values in the final score vector C_{38}. Statistics usually tells us that if we wish to learn about the specific characteristics of a given population that is impossible to measure or compute, we can use statistical inferential methods. That is, based on the values obtained from a random sample of a "reasonable size" from the population under study, we can always extract information that allows us to estimate the characteristics of interest and, thus, be able to generalize the obtained conclusions to the population under study. In this specific case, we can interpret our proposed procedure as one that randomly extracts or samples possible final scores (or standings) among the set of all final scores (or standings) that we have in the First Division Spanish Soccer League. The use of the assumption of equal probabilities for the three possible results R_1, R_2 and R_3 in the simulations guarantees that, in the random extraction or sampling, all possible score or standing vectors will be equally likely or have the same probability of being selected in the sample.

In addition, if we use the *equal strength* assumption with a probability that the result of a draw is $p_2 = 0.25$, and also using M =10,000 replications, we obtain that $\Pr(C_{i38} \leq 44) = 0.1728$ and $\Pr(C_{i38} \leq 43) = 0.1427$.

3.3. Exact probability computation

The specific probability computation under study does not require the use of either the normal approximation or the Monte Carlo simulation approaches proposed in the previous sections. More specifically, if we use enumeration techniques it is possible to find the exact probability distribution for C_{i38}.

In Section 3 we have seen that the random variable $X = (X_1, X_2, X_3)'$ follows a multinomial distribution, with probability mass function given by equation (1). Therefore, as $n = 38$, each one of its individual components, $X_i,\ i = 1,2,3$ will take on values in the set $\{0,1,2,\ldots,38\}$, and the set of possible values for the random variable X is finite (i.e., there are $(n+1)(n+2)/2 = 780$ possible values), so that the probability for each one of its possible values can be easily computed by using (1). Once these probabilities have been obtained, for each one of them, we can compute $C_{i38} = 3X_1 + X_2$ and the probabilities for each one of the $(3n+1) = 115$ possible different values for C_{i38} can be easily added up together. We have done this in `Gretl` and list both the possible values and corresponding probabilities for C_{i38} (see Table 3, column with $p_2 = 1/3$). Table 3 includes the values $\Pr(C_{i38} \leq 43) = 0.1768$ and $\Pr(C_{i38} \leq 42) = 0.1444$, values that are very close to those reported in the approximation methods proposed in Sections 3.1 and 3.2 and, thus, these results confirm the conclusion (see the results reported in Sections 3.1 and 3.2) that a team wishing to stay in the first division should obtain at least 43 points at the end of the regular season. One of the anonymous reviewers suggested an alternative procedure to compute the exact probabilities by means of the probability generating function (pgf) of the random variable X, that measures the number of points gained in a match. That is,

$$g(t) = p_1 + p_2 t + p_3 t^3$$

In this sense, the sum of 38 such random variables has a pgf given by $g(t)^{38}$ and, thus, the probability values associated to each score in the distribution function is given by the coefficients of this polynomial.

The previously reported results have used the equiprobability assumption, something that may be a very restrictive assumption in the view of some researchers. However, as we have already mentioned in Section 3, under the *equal strength* assumption, the distribution of the random variable X is also multinomial. As above, from this resulting distribution and with the use of enumeration methods, we can obtain the exact probability distribution for $C_{i38} = 3X_1 + X_2$ as well. Moreover, we can easily compute the probability distribution function for C_{i38} for different values of p_2 (i.e., the probability that the result of the game is a draw). Figure 1 includes the probability distribution functions for different values of p_2 (i.e., for $p_2 = 0.1,\ldots,0.9$). In addition, Table 3 only includes the probability values for the so called central values of the distribution of the Score variable, also as a function of p_2. As can be seen in Table 3, we have included the corresponding probability values for several cases of the *equal strength* model, which contains two of its special cases we have been studying so far: the equiprobability case

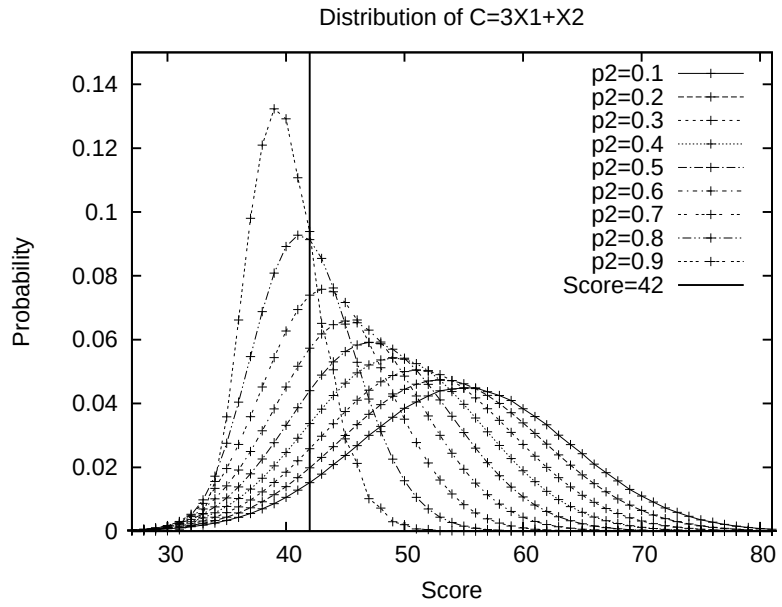


Figure 1: *Team final scores probability distributions as a function of the probability p_2 that the result of the game is a draw.*

Table 3: *Final scores cumulative probability values $C_{i38} = 3X_1 + X_2$ as a function of the probability of a draw, p_2. Reported results correspond to final scores ranging from* 36 *to* 50 *after the regular season has ended and for a twenty-team league. Boldfaced numbers indicate the required final score a team should have at the end of the regular season for not losing the category under the assumed p_2 probability.*

	Probability of a draw: p_2										
Score	0.10	0.20	0.25	0.30	1/3	0.40	0.50	0.60	0.70	0.80	0.90
36	0.0167	0.0217	0.0246	0.0281	0.0306	0.0364	0.0474	0.0617	0.0804	0.1039	0.1252
37	0.0223	0.0290	0.0331	0.0378	0.0414	0.0496	0.0653	0.0867	0.1162	**0.1586**	**0.2232**
38	0.0292	0.0382	0.0437	0.0501	0.0550	0.0661	0.0879	0.1181	**0.1614**	0.2274	0.3442
39	0.0379	0.0496	0.0569	0.0654	0.0717	0.0866	0.1156	**0.1564**	0.2157	0.3082	0.4765
40	0.0484	0.0636	0.0730	0.0838	0.0920	0.1112	0.1488	0.2016	0.2784	0.3974	0.6057
41	0.0613	0.0804	0.0922	0.1059	0.1162	0.1403	**0.1875**	0.2533	0.3479	0.4901	0.7164
42	0.0765	0.1002	0.1148	0.1317	0.1444	**0.1740**	0.2315	0.3106	0.4218	0.5814	0.8103
43	0.0944	0.1233	0.1410	**0.1614**	**0.1768**	0.2122	0.2803	0.3724	0.4976	0.6669	0.8754
44	0.1152	0.1498	**0.1709**	0.1951	0.2132	0.2548	0.3334	0.4370	0.5726	0.7431	0.9260
45	0.1389	**0.1797**	0.2044	0.2326	0.2535	0.3011	0.3896	0.5029	0.6442	0.8080	0.9549
46	**0.1657**	0.2131	0.2415	0.2736	0.2973	0.3507	0.4480	0.5682	0.7103	0.8609	0.9761
47	0.1957	0.2497	0.2818	0.3178	0.3441	0.4028	0.5071	0.6311	0.7694	0.9023	0.9863
48	0.2286	0.2893	0.3250	0.3646	0.3932	0.4564	0.5658	0.6904	0.8207	0.9335	0.9935
49	0.2643	0.3316	0.3705	0.4133	0.4440	0.5106	0.6228	0.7447	0.8637	0.9560	0.9965
50	0.3027	0.3760	0.4178	0.4633	0.4955	0.5645	0.6769	0.7933	0.8987	0.9718	0.9985
51	0.3432	0.4220	0.4663	0.5137	0.5470	0.6170	0.7274	0.8356	0.9265	0.9824	0.9992

(i.e., $p_2 = 1/3$) and the case for which $p_2 = 0.25$. For this latter case, we can clearly see that $\Pr(C_{i38} \leq 44) = 0.1709$ and $\Pr(C_{i38} \leq 43) = 0.1410$, so that, $C_{i38c} = 44$.

4. Probability of not losing the category

In order to compute the probability of not losing the category for a given team having a final score c, we would have to check the joint distribution of the scores for all twenty teams in the league at the end of the regular season and see in how many cases a team having c points has not lost the category. This implies working with the joint distribution of a 20×1 vector of random variables, in which each of its individual components would have a similar distribution to that described for C_{i38} in Section 3.3. In addition, we have to point out that these individual variables (i.e., $C_{1,38}, \ldots, C_{20,38}$) are not independent random variables, which makes this a complicated theoretical problem to solve. However, it is not difficult to obtain an approximation of this distribution by using a Monte Carlo simulation approximation, such as the one previously described in Section 3.2.

In order to describe this new approach, we define a binary random variable D_1, taking value one if the score c appears as part of the final standings score vector C_{38} and if, in addition, it is larger than the score obtained by the team appearing in the eighteenth final standings ordered position vector C^o_{38}, and zero otherwise. That is,

$$D_1 = \begin{cases} 1 & \text{if } c \in C_{38} \text{ and } c > C^o_{18,38} \\ 0 & \text{otherwise} \end{cases}$$

Therefore, in the simulation process described in step 3 of Section 3.2, each time a simulated final score vector C_{38} is obtained, the random variable D_1 takes on two possible values, one or zero. After a sufficiently large number of replications M has been simulated, and if we let m_c be the number of occasions in which the random variable D_1 has taken value one, and M_c be the total number of occasions in which the value c appeared in the final standings score vector C_{38}, then, for a team obtaining c points at the end of the regular season, m_c/M_c would be an approximation of the probability of not losing the category $p_{nlc}(c)$. We should indicate that in a simulation with a finite number of replications we could have two easy to handle types of indetermination showing up: cases with very low scores or cases with very high scores. More specifically, none of the 50,000 replications performed to obtain the results reported in Figure 2 included scores lower than 15 points or higher than 88 points. The problem is solved by assigning probability zero to the very low values and probability one to the very high values obtained in the simulation process. These simulations were carried out for $M = 50,000$, and values of $c = 0, 1, 2, \ldots, 114$ were considered, with the results reported in Figure 2. In this specific case, the probabilities of not losing the category for a team having 42

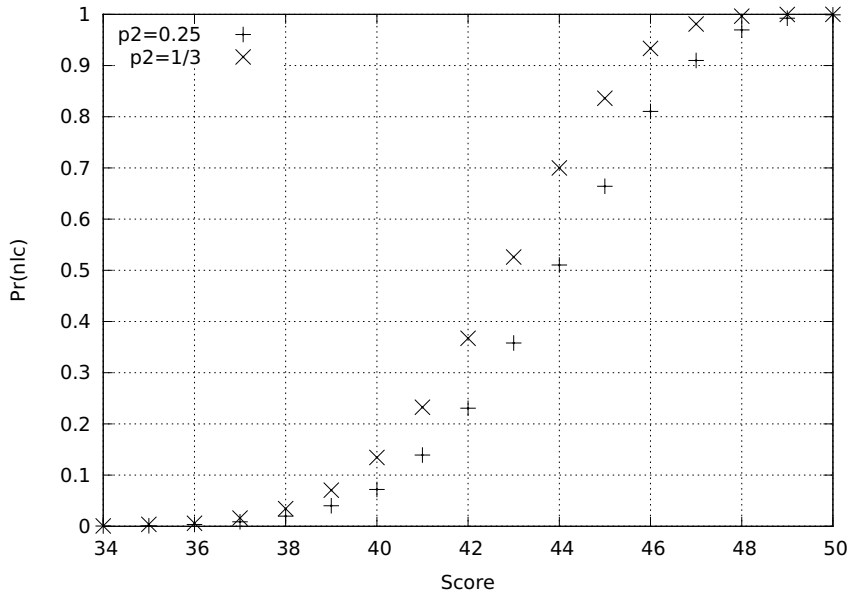


Figure 2: *Probability of not losing the category, p_{nlc} for each final score. Probabilities were computed by simulations and with M =50,000 replications, with $p_1 = p_3$ and two different values for the probability of a draw, $p_2 = 1/3$ (i.e., equiprobability assumption) and $p_2 = 0.25$ (i.e., equal strength assumption).*

and 43 points are 0.3673 and 0.5259, respectively, under the equiprobability assumption, and 0.2307 and 0.3581, respectively, under the *equal strength* assumption with $p_2 = 0.25$. If one wishes to compute instead the probability of playing the European Champions League or the Europa League tournament, the binary variable should be defined accordingly.

5. Dynamic computation of probabilities during the regular season

From a practitioners' point of view, it would be very interesting to have the possibility of computing, after a given number of rounds of games have been played (say k) and conditioned on the current score vector (say C_k), the probability that a given team wins the league, plays the European Champions League or ends in a position that will make that team not to lose its category at the end of the regular season. These probabilities can then be used by the teams to make strategic decisions during the regular season and not at the end of it when things cannot be changed. For example, a team whose main objective at the end of the current regular season is to stay in the first division, would be able to determine that, if the probability of not losing the category, say at mid-season, is smaller than 0.10, the team will change its coach at mid-season. However, another team whose main objective at the end of the regular season is to win the league

could decide that it would change its coach if the probability of winning the league at mid-season is lower that 0.75. That is, conditions and decision are highly linked to the team's objectives during the regular season. We should also mention that, depending on the specific conditions of the league under study, it is very likely that our equiprobability or equal strength assumptions provide a solution that may not be too realistic. In fact, if there are reasons that lead us to believe that these hypotheses do not hold, we should probably propose a more complex or general probability model that allows us to improve the reported results. In any case, we do believe that for any soccer league it would be interesting and useful to have the reference values that can be easily obtained from the model under the aforementioned assumptions.

Therefore, our aim is to be able to compute, for a given team A_i, a given date k for which games were scheduled during the regular season ($k < 38$), and conditioned on the current score vector C_k, the dynamic conditioned probability that at the end of the regular season the team A_i loses its category (plays the European Champions League, plays the Europa League tournament or wins the league).

The method proposed in Section 3.3, in which we now have $n = 38 - k$ can be used to compute the exact probability distribution for each team's score at the end of the regular season, conditioned on the score each team has at the k-th round of games, C_{ik}. That is, we would be able to find the marginal distribution of the random variable C_{i38} (team's A_i score at the end of the regular season), conditioned on the current information we have for the k-th round of games. However, in order to compute the probability that, at the end of the regular season, a given team does not lose its category, conditioned on the current information we have (say C_k), it is necessary to take into account the complete structure the score vector C_k has; that is, the score all twenty teams have at that specific date. From this information, the computation of the probability of a team not losing its category means, as we saw in Section 4 above, working with the joint probability distribution of the scores for all twenty teams. Moreover, if we consider that those scores are not independent we will soon arrive at the conclusion that the analytical computation of this probability is a complicated probability problem, just as we had in Section 4.

As one can see, this is also a very simple problem if we decide to use Monte Carlo simulation techniques to solve it. There are differences, however, with the solution we proposed in Section 4, which will be described in detail below. In this case, the simulation process would start by taking the scores in the k-th date as given or known (i.e., C_k is assumed to be known) and, thus, we would only need to simulate the results for the remaining $38 - k$ dates for which games are scheduled. The whole process can be easily summarized as follows:

- We assume that the 20×1 current score vector for the k-th round of games, C_k, is known.

- For a given team A_i, we define a binary random variable D_2, taking value one if, at the end of the regular season, the team's position in the final standings ordered

score vector C^o_{38} is in one of its first seventeenth places, and zero otherwise. That is, D_2 will take value one if team A_i's score value $c = C_{i38}$ at the end of the regular season is larger that the score obtained by the team at the eighteenth position in the final standings ordered score vector C^o_{38} (i.e., $C^o_{18,38}$), and zero otherwise. We should point out that we are not taking into account any additional criteria such as, for example, goal differences that would decide the final position of two teams (i.e., the ones in positions seventeenth and eighteenth) in case of two teams having the same score, mainly because after thirty-eight games this is not so likely to occur. That is,

$$D_2 = \begin{cases} 1 & \text{if } C_{i38} > C^o_{18,38} \\ 0 & \text{if } C_{i38} \leq C^o_{18,38} \end{cases}$$

which can be easily done simultaneously for all twenty teams, so that we would now have a 20×1 vector of binary indicator variables.

- For the remaining $38-k$ rounds of games, repeat step 3 in the simulation process described in Section 3.2, so that we obtain the final standings ordered score vector C^o_{38} at the end of the regular season. The binary variable D_2 will then take on values one or zero.

- Repeat the whole process of generating the remaining $38-k$ dates for which games are scheduled for a sufficiently large number of replications M. In each replication, the binary variable will take on values one or zero. If we let m be the number of occasions in which the binary variable D_2 has taken value one, then m/M would be an approximation of the probability of team's A_i not losing its category $p_{nlc}(c)$, conditioned on the current score vector C_k.

We now apply this to the soccer league motivating our proposals (see Table 4). The third column in Table 4 (labelled as C^o_{19} in the left-hand side of the table), includes the standings for the First Division Spanish Soccer League after the $k = 19$-th round of games (January 24, 2010). Using the method just described in this section and $M = 10,000$ replications, we have computed the probabilities, conditioned on the scores at $k = 19$, of not losing the category, playing at least the "Europa League" (formerly UEFA tournament), playing the European Champions League, and winning the league for all twenty teams in the 2009-2010 regular season. These results are listed on the right-hand side of Table 4. In order to compare this prediction, based on the information available when about 50% of the regular games were played, with the actual final standings for the last regular season, it is probably worth noting that Barcelona won the league and that, in addition, Real Madrid, Valencia, and Sevilla classified to play the European Champions League. Furthermore, Mallorca and Getafe classified to play the "UEFA Europa League", and Valladolid, Tenerife and Xerez lost their category. There were

Table 4: *Teams' classification in the First Division Spanish Soccer League and probabilities computed with the Monte Carlo simulations approximation. In this case, we were in the $k = 19$-th date for which games were scheduled-January 24, 2010 (for Pr(Win), if two or more teams have the same number of points at the end of the regular season, a tie-breaking mechanism that uses a uniform random variable has been applied).*

	Team	C^o_{19}	p_{nlc}	Pr(Europa)	Pr(Champ)	Pr(Win)
1	Barcelona	49	1.0000	0.9983	0.9899	0.6812
2	Real Madrid	44	1.0000	0.9870	0.9452	0.2318
3	Valencia	39	1.0000	0.9190	0.7570	0.0591
4	Mallorca	34	0.9989	0.7043	0.3838	0.0088
5	Deportivo	34	0.9992	0.6916	0.3758	0.0093
6	Sevilla	33	0.9976	0.6370	0.3068	0.0055
7	Getafe	30	0.9900	0.3876	0.1471	0.0022
8	Athletic	30	0.9900	0.4016	0.1525	0.0022
9	Villarreal	26	0.9399	0.1580	0.0425	0.0001
10	Sporting	24	0.9037	0.0829	0.0175	0.0001
11	Atlético	23	0.8894	0.0671	0.0160	0.0002
12	Osasuna	23	0.8867	0.0653	0.0212	0.0001
13	Rácing	23	0.8825	0.0738	0.0244	0.0001
14	Espanyol	20	0.7389	0.0317	0.0062	0.0000
15	Almería	18	0.5951	0.0148	0.0018	0.0000
16	Málaga	17	0.5120	0.0084	0.0011	0.0000
17	Valladolid	17	0.5234	0.0083	0.0012	0.0000
18	Tenerife	17	0.5089	0.0068	0.0006	0.0000
19	Zaragoza	14	0.3021	0.0009	0.0001	0.0000
20	Xerez	8	0.0689	0.0001	0.0001	0.0000

two relevant issues that provided not expected results for the 2009-2010 regular season: Zaragoza did not lose its category and Deportivo did not play the Europa League. Zaragoza's performance during the second half of the regular season was quite better than that in the first half of the regular season (obtaining 27 points out of 57 possible points), a fact that allowed the team to stay in the first division. Deportivo's performance during the second half of the regular season was quite unexpectedly bad (obtaining only 13 points out of the possible 57 points, while in the first half it had obtained 34 out of 57 points). As can be clearly seen, this is a fact the proposed method clearly did not take into account because its prediction was based on past data. Of course, it is clear that dynamic predictions would be better as we approach the end of the regular season.

6. Conclusions and practical recommendations

We have proposed an approximate method to compute the probability that a team having 42 points has of losing its right to play in the first division the next regular

season. Under the assumption that all possible classifications are equally likely, this method allows us to obtain an estimated value of 0.3673 for this probability, and, an estimated value of 0.2307 under the *equal strength* assumption with probability of a draw of $p_2 = 0.25$.

We have described the normal and Monte Carlo simulated approximations, as well as the exact method, to estimate what would be the objective score a team should aim for in order to stay in the first division of the Spanish soccer League. All three methods have concluded that the objective score for such a team should be of at least 43 points.

Finally, we have also proposed a simulation-based method that allows us to compute, in a dynamic form and after the k-th round of games has ended, the probability, conditioned on the scores it has up to and including that k-th date, of a team not losing its category (or winning the league, of playing the European Champions League or the Europa League tournament)

As we have already mentioned in previous sections, the equiprobability and equal strength assumptions, even after being considered too simplistic or not too realistic hypotheses, have two fundamental and very relevant advantages: under these assumptions, computations are quite simple because of their underlying independence assumption between games, and, in addition, they do not require of any additional a priori information to be able to compute the probabilities of interest. In practice, if one wishes to study the problem of a "real" league, just like the First Division Spanish Soccer League in which there are real reasons to believe that the probability of winning a game, losing a game or that the result of the game is a draw for each team is different (i.e., large or even extreme differences in the budgets for the different teams), then the results reported here can be only considered as upper or lower bounds for the probabilities of interest. For example, it is quite reasonable to believe that a team in the first (or last) position in the league will have a probability larger (or smaller) than $1/3$ of winning most of its games and this can clearly result in the fact that the probability values reported here for winning the league or winning a place to compete in the Champions League (or losing its category) for this specific team can be then considered as lower (or upper) bounds for the real probability of interest.

Future research includes the possibility of not having equally likely classifications or adding some additional information, such as some differential characteristics the different teams in the league have. For example, teams having a larger budget (i.e., richer teams) have more possibilities of bringing better players to their teams. One way of approaching this new problem could be, for example, to establish an *a priori* probability of winning each game that somehow depends on the team's budget. An additional possibility would be to establish this probability taking previous results as the basis for it. In any case, this is out of the scope of this paper and it will be the objective of future research.

Acknowledgements

This research was supported by grants SEJ2007-61362/ECON, MTM2007-60112, ECO 2010-15332 and MTM2010-14913 (Ministerio Español de Ciencia e Innovación and FEDER), and IT-334-07 (Departamento de Educación del Gobierno Vasco - UPV/EHU Econometrics Research Group). The authors would also like to thank two reviewers for providing thoughtful comments and suggestions which led to substantial improvement of the paper.

References

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Brillinger, D. R. (2008). Modelling game outcome of the Brazilian 2006 Series A Championship as ordinal-valued. *Brazilian Journal of Probability and Statistics*, 22, 89-104.

Cottrell, A. and Lucchetti, R. (2009) Gretl User's Guide. Gnu Regression, Econometrics and Time Series. http://sourceforge.net/projects/gretl/files/manual/ [Online; November, 2009 version].

Cryer, J. B. and Miller, R. B. (1991). *Statistics for Business: Data Analysis and Modelling*. Boston: PWS-KENT publishing Company.

Díaz-Emparanza, I. (2002). Is a small Monte Carlo analysis a good analysis? Checking the size, power and consistency of a simulation-based test. *Statistical Papers*, 43(4), 567-577.

Hogg, R. W. and Tanis, E. A. (1988). *Probability and Statistical Inference*. New York: Macmillan Publishing Company.

Karlis D. and Ntzoufras J. (2000). On modelling soccer data. *Student*, 3, 229-245.

Karlis, D. and Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, 20, 133-145

Kleijnen, J. P. C. (1987). *Statistical Tools for Simulation Practitioners*. New York: Marcel Dekker, Inc.

Lee, A. J. (1997). Modeling scores in the premier league: is Manchester United *really* the best? *Chance*, 10, 15-19.

Morris, C. (1975). Central limit theorems for multinomial sums. *The Annals of Statistics*, 14(1), 165-188.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: Wiley.

Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society-Series D (The Statistician)*, 49, 399-418.

SORT 34 (2) July-December 2010, 201-222

Bayes linear spaces

Karl Gerald van den Boogaart[1], Juan José Egozcue[2],
and Vera Pawlowsky-Glahn[3]

Abstract

Linear spaces consisting of σ-finite probability measures and infinite measures (improper priors and likelihood functions) are defined. The commutative group operation, called perturbation, is the updating given by Bayes theorem; the inverse operation is the Radon-Nikodym derivative. Bayes spaces of measures are sets of classes of proportional measures. In this framework, basic notions of mathematical statistics get a simple algebraic interpretation. For example, exponential families appear as affine subspaces with their sufficient statistics as a basis. Bayesian statistics, in particular some well-known properties of conjugated priors and likelihood functions, are revisited and slightly extended.

MSC: 60A10, 62E10

Keywords: Aitchison geometry, compositional data, exponential families, likelihood functions, probability measures, Radon-Nikodym derivative

1. Introduction

More than two decades ago, J. Aitchison (1986) noted that perturbation in the D-part simplex, the sample space of compositional data with a finite number of parts, *"is familiar in other areas of statistics ... as the operation of Bayes's formula to change a prior probability assessment into a posterior probability assessment through the perturbing influence of the likelihood function"* (Aitchison, 1986, p. 45). Recently, the linear space structure of the simplex has been recognised, with perturbation as the Abelian

[1] Institut für Stochastik, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, Prüferstraße 9, D-09596 Freiberg, Germany, Tel.: 0049-(03731) 39-3225, Fax: 0049-(03731) 39-3598, boogaart@math.tu-freiberg.de

[2] Univ. Politécnica de Catalunya, Dep. Matemática Aplicada III, juan.jose.egozcue@upc.edu

[3] Univ. de Girona, Dep. Informática y Matemática Aplicada, vera.pawlowsky@udg.edu

Received: February 2010
Accepted: October 2010

group operation, and its Euclidean structure has been completed (Billheimer *et al.*, 2001; Pawlowsky-Glahn and Egozcue, 2001, 2002; Egozcue *et al.*, 2003) The extension of the underlying ideas to compositions of infinitely many parts is due to Egozcue *et al.* (2006). It leads to the study of probability densities with support on a finite interval, concluding with a Hilbert space structure based on the natural generalisation of the operations between compositions to operations between densities. The space contains both densities corresponding to finite measures, equivalent to probability measures, and densities corresponding to infinite measures, such as likelihood functions or improper (prior) densities. The extension to infinite support measures was suggested as an open problem and is now presented here.

Many different algebraic structures can be defined on sets of positive measures, and particulary on probability measures. For instance, certain classes of measures form a semi-group with respect to the ordinary sum or to the convolution (Bauer, 1992); Markov processes give rise to a semi-group of transition kernels (Markov-semigroups) (Bauer, 1992); $L^p(\lambda)$ can be seen as a space of densities of signed measures; random variables with variance constitute a Hilbert space (Witting, 1985; Small and Leish, 1994; Berlinet and Thomas-Agnan, 2004), which is relevant in statistical modelling; metric spaces are obtained defining distances such as Hellinger-Matusita (Hellinger, 1909; Matusita, 1955) or those based on Fisher-information. Finally, kernel reproducing Hilbert spaces (Whaba, 1990; Berlinet and Thomas-Agnan, 2004) are used for modelling stochastic processes, random measures and nonparametric functions, as well as linear observations of them, the inner product, reproducing kernel, and distance, being related to the variance of the process, and the elements of the space being realisations of stochastic processes (Whaba, 1990).

However, none of the above mentioned structures postulates Bayes updating as a group operation. Bayes theorem has two important characteristics that make it attractive as an operation between measures: (i) it has been considered as a paradigm of information acquisition, and (ii) it is a natural operation between densities (e.g. in probability, Bayesian updating; in system analysis, filtering in the frequency domain).

The primary goal of the present contribution is to provide a linear space structure for sets of classes of densities associated with positive measures of any support. The support of a density is treated as a measure itself, leading to a general and inclusive framework. In particular, linear spaces whose elements are classes of σ-additive positive measures – including probability measures, prior densities and likelihood functions – are introduced. Such spaces are suitable to review many issues of probabilistic modelling and statistics. We call them Bayes spaces because the Abelian group operation, or perturbation for short, corresponds to the operation implied in Bayes theorem. Section 2 defines Bayes linear spaces and Section 3 discusses their affine properties. Exponential families of distributions are identified as affine spaces in Section 4. In Section 5 a review of probabilistic models involved in Bayesian statistics is presented.

2. Bayes linear spaces

Standard tools of measure theory (Ash, 1972; Bauer, 1992, 2002; Shao, 1999) will be useful in the following development. Let λ be a σ-finite, positive measure on an arbitrary measurable space $(\Omega, \mathscr{B})$, where Ω is a non-empty set and $\mathscr{B}$ is a σ-field on Ω. The symbols λ and $\mathscr{B}$ have been chosen deliberately to associate them with the Lebesgue-measure and the Borelian σ-field, as they are a typical example for λ and $\mathscr{B}$. Measures with the same null-sets are called equivalent (Bauer, 1992). This is a very inclusive equivalence relation identifying e.g. the Lebesgue-measure – measuring the volume of a space portion – with any measure with positive density on the same measurable space. The class of measures equivalent to a reference measure, λ, is used to constitute the elements of the Bayes space:

Definition 1 (Equivalent measures) *Let λ and μ be σ-finite measures on $(\Omega, \mathscr{B})$. They are equivalent if, for all $R \in \mathscr{B}$, $\lambda(R) = 0$ if and only if $\mu(R) = 0$. The class of σ-finite measures on $(\Omega, \mathscr{B})$ equivalent to a given reference measure λ is denoted by $\mathscr{M}(\lambda)$ and its elements are called λ-equivalent measures.*

The Radon-Nikodym derivative theorem and the chain rule for densities are stated in the context of equivalent measures. The Radon-Nikodym-derivative is used to identify measures with functions:

Theorem 1 (Radon-Nikodym derivative) *Let λ be a σ-finite measure on $(\Omega, \mathscr{B})$, and μ a σ-finite λ-equivalent measure. Then, there exists a λ-almost-everywhere, $\lambda-$a.e., unique positive function $f : \Omega \to \mathbb{R}_+ = (0, \infty)$ such that, for any $R \in \mathscr{B}$, $\int_R d\mu = \int_R f \, d\lambda$. The function f is then called density, or Radon-Nikodym-derivative, of μ with respect to λ, and is denoted by*

$$\frac{d\mu}{d\lambda}(x) = f(x) \; .$$

Every measure in $\mathscr{M}(\lambda)$ can be represented by a unique density defined $\lambda-$a.e.. The chain rule is closely related to addition and difference in the Bayes linear space:

Theorem 2 (Chain rule for densities) *Let μ, ν be λ-equivalent measures. Then*

$$\frac{d\mu}{d\nu} = \frac{d\mu}{d\lambda} \, \frac{d\lambda}{d\nu} \; .$$

The aim of the following definitions is to build a linear space of classes of σ-finite measures represented either by probability measures or by infinite measures. The first step consists in identifying measures which differ only in a scale factor, leading to equivalence classes of proportional measures. As a consequence, finite measures can be represented by probability measures integrating to one. This idea has been previously used for densities on an interval in (Egozcue *et al.*, 2006) and goes back to a similar

idea which identifies equivalence classes of positive vectors with compositions (Barceló-Vidal *et al.*, 2001).

Definition 2 (*B*-equivalence) *Let μ and ν be measures in $\mathscr{M}(\lambda)$. They are B-equivalent, $\mu =_B \nu$, if and only if there exists a constant $c > 0$ such that, for any $R \in \mathscr{B}$, $\mu(R) = c \cdot \nu(R)$, using the convention $c \cdot (+\infty) = +\infty$. The set of $(=_B)$ equivalent classes is denoted as a quotient space $B(\lambda) = \mathscr{M}(\lambda)/(=_B)$.*

Theorem 3 *$(=_B)$ is an equivalence relation on $\mathscr{M}(\lambda)$.*

The elements of $B(\lambda) = \mathscr{M}(\lambda)/(=_B)$ are $(=_B)$-equivalence classes of measures in $\mathscr{M}(\lambda)$. From now on, no notational difference will be made between a measure and the equivalence class it represents. When a reference measure λ is fixed, a $(=_B)$-class of measures will be represented by a density (or Radon-Nikodym derivative with respect to λ) defined $\lambda-$a.e. and up to a positive constant. The equivalence symbol $(=_B)$ will be used for $\mu, \nu \in \mathscr{M}(\lambda)$ and for their respective densities, f_μ and f_ν. Thus, if $\mu =_B \nu$, then $f_\nu =_B f_\mu$, which means that there exists c such that $f_\nu(x) = c f_\mu(x)$ $\lambda-$a.e. Summarising, $(=_B)$ identifies a measure equivalence class with a density, and the measures are all seen as the same element of $B(\lambda)$. To build a linear space on $B(\lambda)$, the second step consists in introducing addition and multiplication by real scalars.

Definition 3 (Perturbation and powering) *Let μ and ν be measures in $B(\lambda)$. For every $R \in \mathscr{B}$, the perturbation of μ by ν is the measure in $B(\lambda)$ such that*

$$(\mu \oplus \nu)(R) = \int_R \frac{d\mu}{d\lambda} \frac{d\nu}{d\lambda} \, d\lambda \, . \tag{1}$$

For a scalar $\alpha \in \mathbb{R}$, the powering of μ is the measure in $B(\lambda)$ such that

$$(\alpha \odot \mu)(R) = \int_R \left(\frac{d\mu}{d\lambda} \right)^\alpha d\lambda \tag{2}$$

Theorem 4 *Perturbation and powering of σ-finite λ-equivalent measures are σ-finite.*

Proof: see appendix.

Perturbation and powering of σ-finite λ-equivalent measures are based on perturbation and powering in the simplex, introduced originally by J. Aitchison (Aitchison, 1986) and shown later to structure the simplex as a linear space (Martín-Fernández *et al.*, 1999; Billheimer *et al.*, 2001; Pawlowsky-Glahn and Egozcue, 2001; Aitchison *et al.*, 2002). The space is denoted $B(\lambda)$ (B for Bayes) recalling that perturbation, which plays the role of group operation, is essentially the operation in Bayes theorem.

The inverse operation of perturbation in $B(\lambda)$, i.e. substraction in $B(\lambda)$, is defined as $\ominus\mu =_B (-1) \odot \mu$. The use of densities representing the corresponding measures

generates alternative definitions of perturbation and powering. Let f_μ and f_ν be densities in $B(\lambda)$ and $\alpha \in \mathbb{R}$; then, perturbation, difference and powering are

$$(f_\nu \oplus f_\mu)(x) =_B f_\nu(x) f_\mu(x)\,, \tag{3}$$

$$(f_\nu \ominus f_\mu)(x) =_B \frac{f_\nu(x)}{f_\mu(x)}\,, \tag{4}$$

$$(\alpha \odot f_\nu)(x) =_B f_\nu(x)^\alpha\,. \tag{5}$$

Combining measures and densities we get equivalent expressions:

$$(f_\nu \oplus \mu) =_B \int_A f_\nu(x) d\mu(x)\,, \tag{6}$$

$$(\nu \ominus \mu)(x) =_B \frac{d\nu}{d\mu}\,. \tag{7}$$

A remarkable fact is that the difference (4), (7) is actually a Radon-Nikodym derivative due to the chain rule (Theorem 2).

When using densities representing measures, operations depend on the reference measure λ adopted. Therefore, whenever not clear from the context, a subscript will be used: $\oplus_\lambda$, $\ominus_\lambda$, $\odot_\lambda$, $=_{B(\lambda)}$.

Theorem 5 *With operations $\oplus$ and $\odot$, $B(\lambda)$ is a real vector space.*

Proof: see appendix.

Whatever the reference measure λ, the neutral element of $B(\lambda)$ with respect to perturbation is a constant density, or equivalently, the density with constant value 1. The perturbation-opposite of a density f_μ is B-equivalent to $1/f_\mu$.

Definition 4 (Bayes space) *The linear space $(B(\lambda), \oplus, \odot)$ is called Bayes space with reference measure λ.*

When the measurable space is $(\Omega, \mathscr{B}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\mathscr{B}(\mathbb{R})$ the Borel σ-field on $\mathbb{R}$, the most commonly used reference measure is the Lebesgue measure $\lambda_\mathbb{R}$. For constrained measurable spaces such as the positive real line, $\Omega = \mathbb{R}_+$, or the 3-part simplex, $\Omega = \mathscr{S}^3$, with the corresponding restricted Borelians, the Lebesgue measure restricted to them, λ_+, respectively $\lambda_{\mathscr{S}^3}$, may be readily used. These contexts are usual in probability theory and do not need further examples. Similarly, the measurable spaces of the integers or the non-negative integers, $(\mathbb{Z}, \mathbb{Z}_+)$, are normally used with the counting

measure as a reference. However, different but useful reference measures can be taken in $\mathbb{R}_+$ and in $\mathscr{S}^3$. As they are seldom used, they are given as examples.

Example 1 Consider $(\Omega, \mathscr{B}) = (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$, being $\mathbb{R}_+$ the strictly positive real numbers. A natural reference is the relative measure, defined for any interval $[a,b] \subset \mathbb{R}_+$, as $\mu_+([a,b]) = \ln b - \ln a$, whose density with respect to λ_+ is

$$\frac{d\mu_+}{d\lambda_+} = \frac{d\ \ln(x)}{dx} = \frac{1}{x}\,.$$

The reference measure μ_+ corresponds to a constant density in the space $B(\mu_+)$. Moreover, in $B(\mu_+)$, the density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \xi)^2}{2\sigma^2}\right), \tag{8}$$

represents a log-normal probability law with median $\exp(\xi)$ and logarithmic variance σ^2. It has been called the normal in $\mathbb{R}_+$ (Eaton, 1983; Mateu-Figueras *et al.*, 2002) and is accordingly denoted by $\mathscr{N}_+(\xi, \sigma^2)$. The positive real line, $\mathbb{R}_+$, can be structured as an Euclidean space taking into account that $\ln : \mathbb{R}_+ \to \mathbb{R}$ is a one-to-one mapping (Pawlowsky-Glahn and Egozcue, 2001). Then, μ_+ is induced by the Lebesgue measure in $\mathbb{R}$. Thus, the reference measure μ_+ corresponds to a relative scale in $\mathbb{R}_+$.

Example 2 The unit 3-part simplex, $\mathscr{S}^3 \subset \mathbb{R}^3$, has elements which are vectors with 3 strictly positive components adding to 1. The simplex $\mathscr{S}^3$ has been shown to be a 2-dimensional Euclidean space using perturbation and powering (as operations of its elements) and the Aitchison metrics (Pawlowsky-Glahn and Egozcue, 2001; Billheimer *et al.*, 2001). Consequently, an orthonormal basis can be defined such that elements in the simplex can be represented by the corresponding coordinates. Once an orthonormal basis has been selected, the mapping assigning coordinates to each element of the simplex has been called isometric log-ratio transformation (ilr) (Egozcue *et al.*, 2003). A particular case of ilr can be used to define a new reference measure in $\mathscr{S}^3$ in the following way. Take $\Omega = \mathscr{S}^3$ and consider the one-to-one mapping $\mathrm{ilr} : \mathscr{S}^3 \to \mathbb{R}^2$ defined by

$$\mathrm{ilr}(\vec{x}) = \left(\frac{1}{\sqrt{2}}\ \ln\frac{x_1}{x_2}\,,\ \frac{1}{\sqrt{6}}\ \ln\frac{x_1 x_2}{x_3^2}\right),$$

where $\vec{x} = (x_1, x_2, x_3) \in \mathscr{S}^3$. Define the σ-field $\mathscr{B}(\mathscr{S}^3) = \mathrm{ilr}^{-1}(\mathscr{B}(\mathbb{R}^2))$ and a reference measure $\alpha_{\mathscr{S}^3}(\mathrm{ilr}^{-1}(R)) = \lambda_{\mathbb{R}^2}(R)$, for $R \in \mathscr{B}(\mathbb{R}^2)$. The measure $\alpha_{\mathscr{S}^3}$ is called Aitchison measure (Egozcue *et al.*, 2003; Mateu-Figueras *et al.*, 2003; Pawlowsky-Glahn, 2003). In this context, the additive logistic normal probability distribution (aln) (Aitchison,

1986) is represented by the density

$$f(\vec{x}) = \frac{1}{2\pi|\Sigma|}\exp\left(-\frac{1}{2}(\mathrm{ilr}(\vec{x})-\boldsymbol{\mu})^t\Sigma^{-1}(\mathrm{ilr}(\vec{x})-\boldsymbol{\mu})\right), \tag{9}$$

where vectors of three components in $\mathscr{S}^3$ are denoted using $(\vec{\ })$ and vectors in $\mathbb{R}^2$ are boldfaced. Σ is a $(2,2)$-covariance matrix, $|\Sigma|$ denotes its determinant, and $\boldsymbol{\mu}\in\mathbb{R}^2$ plays the role of a mean because $\mathrm{ilr}^{-1}(\boldsymbol{\mu})$ is actually the centre of the distribution. This probability measure corresponds to Aitchison's aln-probability law or logistic-normal distribution. However, the density (9) has been called normal in $\mathscr{S}^3$ (Mateu-Figueras *et al.*, 2003) because of the absence of the Jacobian of the ilr transformation, which is the density of the reference measure $\alpha_{\mathscr{S}^3}$ with respect to $\lambda_{\mathbb{R}^3}$.

3. Affine transformation and subsets of $B(\lambda)$

Changing the reference measure of $B(\lambda)$ to a B-equivalent one does not change the space. The transformation from $B(\lambda)$ to $B(\mu)$, being $\mu\in\mathscr{M}(\lambda)$, is an affine transformation and may be interpreted as a change of origin.

Theorem 6 *Let μ be a measure in $\mathscr{M}(\lambda)$. Then, $\mu =_B \lambda$ if and only if $B(\mu)$ and $B(\lambda)$ are equal as linear spaces.*

Proof: see appendix.

When changing the reference measure, or the origin, of the space $B(\lambda)$, the identification of density and measure is broken. Next theorem on change of origin is formulated in terms of measures, thus avoiding notation with densities.

Theorem 7 (Change of origin) *For all $\mu\in\mathscr{M}(\lambda)$ the spaces $B(\mu)$ and $B(\lambda)$ have the same elements and are equivalent as affine spaces. Consequently, changing the reference measure is a simple shift operation.*

Proof: see appendix.

In analytic geometry the elements of a linear space can be seen from two different points of view: points in the space and vectors or arrows. The first corresponds to affine geometry, the second to the vector space. In the present context, the elements of $B(\lambda)$ can be represented by measures, e.g. μ, ν. This representation by measures corresponds to *points*. Alternatively, the difference $\mu\ominus\nu =_B d\mu/d\nu$, which is actually a density, correspond to a *vector*, i.e. the difference between points is a *vector*. However, as in analytical geometry, there is no mathematical difference between *points* and *vectors* of any kind. The only practical difference arises when shifting the origin from λ to λ'. The vector representation $d\mu/d\lambda\in B(\lambda)$ of the *point* μ is then shifted by subtracting the new

origin represented as a *vector*: $(d\mu/d\lambda') = (d\mu/d\lambda)(d\lambda/d\lambda') =_B (d\mu/d\lambda) \ominus (d\lambda'/d\lambda)$. Therefore, the use of the density notation $f_\mu = d\mu/d\lambda$ makes sense only when the reference measure λ is clearly specified, because the density changes under change of origin.

The space $B(\lambda)$ contains $(=_B)$-classes of finite measures and other classes of infinite measures (σ-finite). A finite measure μ, can be represented by a probability measure $\mu/\mu(\Omega)$, being $\mu =_B \mu/\mu(\Omega)$. Infinite measures cannot be normalised in this way because the measure of the whole space Ω is then infinite. The latter $(=_B)$-classes contain measures like improper priors or improper likelihood functions appearing regularly in Bayesian statistics. In this context, $(=_B)$-equivalence achieves its full meaning as the likelihood principle that identifies proportional proper or improper densities (Birnbaum, 1962; Leonard and Hsu, 1999; Robert, 2001). This means that the space $B(\lambda)$ is decomposed into two well defined subsets: the set of classes of finite measures, $B_P(\lambda)$ containing proper probability measures; and $B_I(\lambda)$ containing classes of infinite measures. By definition $B_P(\lambda)$ and $B_I(\lambda)$ constitute a partition of $B(\lambda)$. The different role that proper and improper densities play in statistics motivates the following properties concerning $B_P(\lambda)$ and $B_I(\lambda)$. Some properties are related to other two important subsets of $B(\lambda)$, namely the set of measures whose density is upper bounded $\lambda-$a.e., $B_u(\lambda)$, and the set of measures whose densities are double bounded, i.e. such that if f a density in $B(\lambda)$, then there exist a positive constant, b, such that $0 < 1/b < f < b < +\infty$ ($\lambda-$a.e.); this subset is denoted by $B_b(\lambda)$.

Theorem 8

1. $B_P(\lambda)$, $B_I(\lambda)$ *is a partition of* $B(\lambda)$.
2. $B_P(\lambda)$ *is convex.*
3. $B_b(\lambda)$ *is a subspace of* $B(\lambda)$.
4. $B_u(\lambda)$ *is a convex cone.*
5. $B_P(\lambda) \oplus B_u(\lambda) = B_P(\lambda)$.
6. $B_I(\lambda) \ominus B_u(\lambda) = B_I(\lambda)$.
7. $\mu \in B(\lambda)$ *if and only if* $B_P(\mu) = B_P(\lambda)$ *as sets of measures.*
8. $\mu \in B_b(\lambda)$ *if and only if* $B_b(\mu) = B_b(\lambda)$ *as sets of measures.*
9. $\mu \in B_P(\mu)$ *if and only if* $B_b(\mu) \subset B_P(\mu)$.
10. $\mu \in B_I(\mu)$ *if and only if* $B_b(\mu) \subset B_I(\mu)$.

Proof: see appendix.

4. Exponential families as affine spaces

Many commonly used distribution families, including multinomial, normal, beta, gamma and Poisson, are exponential families. A common general definition can be given as follows (Witting, 1985):

Definition 5 (Exponential family) *For λ a measure on a measurable space $(\Omega,\mathscr{B})$, consider a strictly positive measurable function $g:(\Omega,\mathscr{B})\to(\mathbb{R}^+,\mathscr{B}(\mathbb{R})|_{\mathbb{R}^+})$; a vector of measurable functions $\vec{T}=(T_1,T_2,\ldots,T_k)$ with $T_i:(\Omega,\mathscr{B})\to(\mathbb{R},\mathscr{B}(\mathbb{R}))$, $i=1,\ldots,k$; and a function $\vec{\boldsymbol{\theta}}=(\theta_1,\theta_2,\ldots,\theta_k)$, where $\theta_i:A\to\mathbb{R}$ and A is a parameter space. A k-parametric exponential family of distributions, $P_{\vec{\boldsymbol{\alpha}}}$, $\vec{\boldsymbol{\alpha}}\in A$, on $(\Omega,\mathscr{B})$ is given by*

$$\frac{dP_{\vec{\boldsymbol{\alpha}}}}{d\lambda}(x)=f_{\vec{\boldsymbol{\alpha}}}(x)=C(\vec{\boldsymbol{\alpha}})\cdot g(x)\cdot\exp\left[\sum_{j=1}^{k}\theta_j(\vec{\boldsymbol{\alpha}})T_j(x)\right],$$

with a normalisation constant

$$C(\vec{\boldsymbol{\alpha}})=\left(\int\exp\left[\sum_{j=1}^{k}\theta_j(\vec{\boldsymbol{\alpha}})T_j(x)\right]\,g(x)\,d\lambda(x)\right)^{-1}. \tag{10}$$

The exponential family is denoted $\mathrm{Exp}(\lambda,g,\vec{T},\vec{\boldsymbol{\theta}})$. *If k is minimal, the family is called strictly k-parametric.*

The function $\kappa(\vec{\boldsymbol{\alpha}})=-\ln C(\vec{\boldsymbol{\alpha}})$ is called the cumulant function of the family. Classically, the parameter space A is restricted to values of $\vec{\boldsymbol{\alpha}}$ for which $C(\vec{\boldsymbol{\alpha}})$ exists. Frequently, λ is called reference measure, and is typically the Lebesgue measure on $\mathbb{R}$ when the support of the random variable is $\mathbb{R}$, or a counting measure when the support is discrete; $\vec{T}(x)$ defines a set of statistics; and $\vec{\boldsymbol{\theta}}(\vec{\boldsymbol{\alpha}})$ is a mapping of the used parameters $\vec{\boldsymbol{\alpha}}\in A$ into the so-called natural parameters, $\theta_i(\vec{\boldsymbol{\alpha}})$, of the family. The normal family of distributions is a typical case: $g(x)$ is constant; $\vec{T}(x)=(x,x^2)$; $\vec{\boldsymbol{\alpha}}=(m,\sigma^2)$, where m is the mean and σ^2 is the variance; and $\vec{\boldsymbol{\theta}}(\vec{\boldsymbol{\alpha}})=(\theta_1(\vec{\boldsymbol{\alpha}}),\theta_2(\vec{\boldsymbol{\alpha}}))=(m/\sigma^2,-1/(2\sigma^2))$.

As mentioned, classical exponential families are defined only for those $\vec{\boldsymbol{\alpha}}$ for which $\kappa(\vec{\boldsymbol{\alpha}})$ or $C(\vec{\boldsymbol{\alpha}})$ in (10) exists. However, the idea of Bayes spaces permits to drop this condition and infinite measures can be considered natural members of exponential families. A definition of such extended exponential families is the following.

Definition 6 (Extended exponential family) *Using the notation in definition 5, an extended exponential family, denoted* $\mathrm{Exp}_B(\lambda, g, \vec{T}, \vec{\boldsymbol{\theta}})$, *contains the densities*

$$\frac{dP_{\vec{\boldsymbol{\alpha}}}}{d\lambda}(x)=_B f_{\vec{\boldsymbol{\alpha}}}(x)=_B g(x)\cdot\exp\left[\sum_{j=1}^{k}\theta_j(\vec{\boldsymbol{\alpha}})T_j(x)\right].$$

If k is minimal, the family is called strictly k-parametric.

Densities in the extended family may or may not correspond to probability measures. Particularly, the elements with finite integral form the exponential family in the ordinary sense. Next theorems account for the properties of the extended exponential families.

Theorem 9 *An extended exponential family* $\mathrm{Exp}_B(\lambda, g, \vec{T}, \vec{\boldsymbol{\theta}})$ *is a finite dimensional affine subspace of the Bayes space* $B(\lambda)$.

Proof: see appendix.

Theorem 10 *Any k-dimensional affine subspace S of* $B(\lambda)$ *is a strictly k-parametric extended exponential family.*

Proof: see appendix.

When an extended exponential family is viewed as an affine space, g can be identified as the origin of the affine space. Also, the change of origin of $B(\lambda)$ from λ to $\mu =_B \lambda \oplus g$, where g is taken as a density of a σ-finite measure, transforms the exponential family into a subspace of $B(\mu)$ because the constant density or neutral element for $\oplus$ is now an element of the family. Another important aspect is that the natural parameters $\theta_j(\vec{\boldsymbol{\alpha}})$ are the coordinates of $\mu_{\vec{\boldsymbol{\alpha}}}$ expressed in the basis elements $V_j(x)$. The restriction of the parameter space of exponential families, due to the integrability condition for the existence of the normalisation constant, is not any more needed in this context. Non integrable elements correspond to densities of infinite measures in $B_I(\lambda)$. When exponential families must be used as families of probability distributions, improper distributions can be just ignored and restrictions to the parameters apply.

Example 3 For $\Omega = \mathbb{R}_+$, and using the notation of Example 1, the log-normal exponential family is

$$f_{\xi,v}(x) = \frac{dP_{\xi,v}}{d\lambda_+}(x) = \frac{1}{\sqrt{2\pi v}} \cdot \frac{1}{x} \cdot \exp\left(-\frac{(\ln x - \xi)^2}{2v}\right),$$

where v is the logarithmic variance and $C(\xi, v) = \exp(-\frac{1}{2v}\xi^2)/\sqrt{(2\pi v)}$, $g(x) = 1/x$, $\vec{\boldsymbol{\theta}} = (\xi/v, -1/(2v))$ and $\vec{T} = (\ln x, (\ln x)^2)$. However, for real values of ξ and positive values of v, $\theta_2 = -1/(2v) < 0$; this means that the family only spans half of the affine space, an affine cone, in $B(\lambda_+)$. The whole affine space is spanned accepting values $v < 0$; for these values, $f_{\xi,v}(x)$ is no longer a probability density but it belongs to $B_I(\lambda_+) \subset B(\lambda_+)$. Additionally, changing the origin from λ_+ to $\mu_+ =_B 1/x$ the family adopts the form

$$\frac{dP_{\xi,v}}{d\mu_+} =_{B(\mu_+)} \exp\left(-\frac{(\ln x - \xi)^2}{2v}\right),$$

which is again the normal in $\mathbb{R}_+$ (8) given in Example 1. The family can be expressed as a subspace of $B(\mu_+)$,

$$\frac{dP_{\xi,v}}{d\mu_+} =_{B(\mu_+)} \left(\frac{\xi}{v} \odot e^x\right) \oplus \left(\frac{1}{v} \odot \frac{dP_{0,1}}{d\mu_+}\right),$$

whereas the family span is an affine subspace of $B(\lambda_+)$,

$$\frac{dP_{\xi,\nu}}{d\lambda_+} =_{B(\lambda_+)} \frac{1}{x} \oplus \left(\frac{\xi}{\nu} \odot e^x\right) \oplus \left(\frac{1}{\nu} \odot \frac{dP_{0,1}}{d\lambda_+}\right) .$$

5. Bayes theorem is summing information

The following context is inspired by Bayesian statistics, however it is also relevant in likelihood function based statistics. For the observations consider a measurable space $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, with $\mathscr{B}(\mathscr{X})$ a σ-field on $\mathscr{X}$, and a reference measure on it denoted by λ. Let $\vec{x} = (x_1, x_2, \ldots, x_n) \in \mathscr{X}^n$ be the vector of observations modelled by independent random variables X_i with values in $\mathscr{X}$ and probability law given by the measure $P_\theta \in B_P(\lambda)$, distribution for short, depending on a set of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ with values in a measurable space $(\Theta, \mathscr{B}(\Theta))$ of parameters. Denote by P_{prior} a prior distribution on $(\Theta, \mathscr{B}(\Theta))$, by P_{post} the posterior, by

$$L_{x_i}(\theta) = \frac{dP_\theta}{d\lambda}(x_i) ,$$

the individual likelihood functions, and by $L_{\vec{x}}(\theta) = \prod_i L_{x_i}(\theta)$ the joint likelihood function. According to the likelihood principle (Leonard and Hsu, 1999), a likelihood L_{x_i} and its scaled version αL_{x_i} should give the same result in the analysis. Thus $=_B$ for functions of θ is a natural equivalence relation for likelihood functions. Consider a reference measure $\tau \in \mathscr{M}(P_{prior})$ on $(\Theta, \mathscr{B}(\Theta))$. Now, two different Bayes spaces are relevant in this situation:

- The Bayes space $B(\lambda)$ containing the family $\{P_\theta : \theta \in \Theta\}$ of distributions for the observations, being $P_\theta \in \mathscr{M}(\lambda)$ on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$.
- The Bayes space $B(\tau)$ containing the distributions of the parameters P_{prior}, P_{post}, for the reference measure τ on $(\Theta, \mathscr{B}(\Theta))$.

Theorem 11 *If the distributions of the family $\{P_\theta : \theta \in \Theta\}$ are in $B(\lambda)$, then $L_{x_i} \in B(\tau)$, $P_\theta - a.e.$*

Proof: see appendix.

In this context, the Bayes formula can be written

$$\frac{dP_{post}}{d\tau}(\theta) = \frac{\frac{dP_{prior}}{d\tau}(\theta) \prod_{i=1}^n L_{x_i}(\theta)}{\int_\Theta \frac{dP_{prior}(\theta)}{d\tau(\theta)} \prod_{i=1}^n L_{x_i}(\theta) d\tau(\theta)} .$$

The denominator is a constant not depending on θ, accordingly,

$$\frac{dP_{post}}{d\tau}(\theta) =_B \frac{dP_{prior}}{d\tau}(\theta) \prod_{i=1}^{n} L_{x_i}(\theta),$$

which, using Bayes space operations, simplifies to the following theorem:

Theorem 12 (Bayes theorem in terms of Bayes spaces) *If $P_\theta \in B(\lambda)$ and the prior $P_{prior} \in B(\tau)$ then, $\bigotimes_{i=1}^{n} P_\theta(x_i)$-a.e.,*

$$P_{post} =_B P_{prior} \oplus \bigoplus_{i=1}^{n} L_{x_i} \tag{11}$$

Bayes theorem has several well-known and interesting direct implications. Here, Theorem 12 is an elegant form of Bayes formula: it is a sum in a vector space and, consequently, Bayesian updating is associative, commutative, invertible and has a neutral element (the non-informative experiment here represented by the measure τ). Also, the addition of the prior is invertible, as the prior can be subtracted and another prior can be added. Thus, adding information in terms of Bayes statistics is nothing but summing vectors in a space of information, here represented by $B(\tau)$. This means that the three densities P_{prior}, $L_{\vec{x}}$ and P_{post} represent information: before the experiment, provided by the experiment, and updated from the experiment respectively. Furthermore, Bayes formula as expressed in Theorem 12, admits both proper or improper priors and improper intermediate posteriors. Also the likelihood function of a repeated independent observation takes the form of a sum:

Corollary 1 *In the conditions of Theorem 12,*

$$L_{\vec{x}} =_B \bigoplus_{i=1}^{n} L_{x_i}$$

6. Bayes theorem and exponential families

In order to simplify the notation, the natural parameters of an exponential family will be used instead of the dependence on general parameters $\vec{\boldsymbol{\theta}}(\vec{\boldsymbol{\alpha}})$; then, arguments of functions of the parameters will be expressed simply as $\vec{\boldsymbol{\theta}}$. The components of the boldfaced vectors of parameters, statistics and observations, are denoted with the same text letters subscripted to indicate component.

Theorem 13 *Let x_i, $i = 1,\ldots,n$, be repeated independent observations from a strictly k parametric exponential family* $\mathrm{Exp}_{B(\lambda)}(\lambda, g, \vec{\boldsymbol{\theta}}, \vec{T})$,

$$P_{\vec{\boldsymbol{\theta}}}(x) = C(\vec{\boldsymbol{\theta}}) \cdot g(x) \cdot \exp\left(\sum_{j=1}^{k} \theta_j T_j(x)\right),$$

then, the joint likelihood $L_{\vec{x}}(\vec{\boldsymbol{\theta}})$, as a function of $\vec{\boldsymbol{\theta}}$, is a $k+1$-parametric family $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$, *with* $g^*(\vec{\theta}) = 1$, $\vec{\boldsymbol{\theta}}^* = (\ln C(\vec{\boldsymbol{\theta}}), \vec{\boldsymbol{\theta}})$, *and* $\vec{T}^*(\vec{x}) = (n, \sum_{i=1}^{n} \vec{T}(x_i))$. *The family is strictly k_1-parametric with $k \le k_1 \le k+1$.*

Proof: see appendix.

A remarkable fact is that the initial statistic $\vec{T}$ plays the role of the vector of natural parameters, $\vec{T}^*$, in the resulting exponential family. Also, note that the first element in $\vec{\boldsymbol{\theta}}^*$ is the negative cumulant function $\kappa(\vec{\boldsymbol{\theta}}) = -\ln C(\vec{\boldsymbol{\theta}})$. Theorem 13 allows the identification of conjugated families of priors and densities of observations.

Theorem 14 *In the conditions of Theorem 13, a prior density $P_{prior}(\vec{\boldsymbol{\theta}})$ in* $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$ *generates a posterior density through the Bayes theorem*

$$P_{post} =_{B(\tau)} L_{\vec{x}} \oplus P_{prior},$$

which is also in $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$, *i.e.* $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$ *and* $\mathrm{Exp}_{B(\lambda)}(\lambda, g, \vec{\boldsymbol{\theta}}, \vec{T})$ *are conjugated families.*

Proof: see appendix.

It is well known that, for exponential families of densities of observations, an exponential family of conjugated priors exists such that it also contains the posteriors (Leonard and Hsu, 1999). Next theorem goes a little bit further, stating that, regardless of the prior, the possible posterior densities are in an extended exponential family.

Theorem 15 *If the likelihood function of a multiple observation $L_{\vec{x}}$ satisfies the conditions of Theorem 13, for any prior $P_{prior} \in B(\tau)$, the posterior, $P_{post} =_{B(\tau)} L_{\vec{x}} \oplus P_{prior}$, is in* $\mathrm{Exp}_{B(\tau)}(\tau, P_{prior}(\vec{\boldsymbol{\theta}}), \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$.

Proof: see appendix.

Next theorem is also a new result for exponential families of posteriors stating the converse of Theorem 15.

Theorem 16 *Assume that the posterior density is obtained from the Bayes formula $P_{post}(\vec{\boldsymbol{\theta}}) =_{B(\tau)} L_{\vec{x}}(\vec{\boldsymbol{\theta}}) \oplus_{\tau} P_{prior}(\vec{\boldsymbol{\theta}})$, where $P_{prior}(\vec{\boldsymbol{\theta}})$ is the prior and the likelihood function is $L_{\vec{x}}(\vec{\boldsymbol{\theta}}) = \prod_{i=1}^{n} L_{x_i}(\vec{\boldsymbol{\theta}})$. If $P_{post}(\vec{\boldsymbol{\theta}}|\vec{x}) \in \mathrm{Exp}_{B(\tau)}(\tau, h, \vec{\boldsymbol{\theta}}, \vec{S})$, then $L_x(\vec{\boldsymbol{\theta}})$, as a func-*

tion of x, is in $\mathrm{Exp}_{B(\lambda)}(\lambda,1,\vec{T},\vec{\boldsymbol{\theta}})$*, for some statistic* $\vec{T}(x)$*. If* $\mathrm{Exp}_{B(\tau)}(\tau,h,\vec{\boldsymbol{\theta}},\vec{S})$ *is k-dimensional, then* $\mathrm{Exp}_{B(\lambda)}(\lambda,1,\vec{T},\vec{\boldsymbol{\theta}})$ *is* k_1*-dimensional with* $k_1 \leq k$*.*

Proof: see appendix.

Corollary 2 *A family of* λ*-equivalent distributions is in an exponential family if and only if, for any prior, the family of its posteriors (perturbation of prior and a member of the family) is an extended exponential family.*

Example 4 Consider $\mathbb{Z}_+$, the non-negative integers, as space of observations, and the counting measure ν as a reference measure on it, i.e. $\nu(\{x\}) = 1$ for any single point $\{x\}$ in $\mathbb{Z}_+$. Define the two-parametric exponential family

$$\mathrm{Exp}(\nu, g(x), (\theta_1,\theta_2), (T_1(x), T_2(x)))\,,$$

with $g(x) = (x!)^{-1}$, $\theta_1 = \ln\phi$, $T_1 = x$, $T_2 = \delta(x)$, with $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. A density of this exponential family has the expression

$$f(x|\phi,\theta_2) = C(\phi,\theta_2)\cdot\frac{1}{x!}\cdot\exp\left(x\ln\phi + \delta(x)\theta_2\right),\ \phi > 0\,, \tag{12}$$

being the normalising constant

$$C(\phi,\theta_2) = \frac{1}{\exp(\theta_2)+\exp(\phi)-1}\,.$$

The density (12) is the Bayes-perturbation in $B(\nu)$ of a Poisson density of parameter ϕ by a step-density $\exp(\theta_2\delta(x))$, the latter in $B_I(\nu)$. However, the whole family is in $B_P(\nu)$ according to Theorem 8, number 5. Note that, for $\theta_2 = 0$, the family reduces to the standard Poisson exponential family. The exponential family (12) may be called *zero-inflated Poisson* family (Lambert, 1992) because it can be written

$$f(x|\phi,\theta_2) = (1-p)\cdot\delta(x) + p\cdot\frac{\phi^x e^{-\phi}}{x!}\,,\ \theta_2 = \ln\left[(1-p)e^{\phi}+p\right]\,,$$

as a mixture of a Dirac and a Poisson distributions, although from the latter expression it is difficult to deduce its exponential character. This zero-inflated Poisson family can also be expressed as an affine subspace of $B(\nu)$

$$f(x|\phi,\theta_2) \;=_{B(\nu)}\; \frac{1}{x!}\oplus(\ln\phi\odot e^x)\oplus\left(\theta_2\odot e^{\delta(x)}\right)\,,$$

or, alternatively, taking $\mu = \nu\ominus(1/x!)$ as reference measure, the family is a subspace of $B(\mu)$

$$f(x|\phi,\theta_2) =_{B(\mu)} (\ln\phi \odot e^x) \oplus \left(\theta_2 \odot e^{\delta(x)}\right) .$$

In both cases, with $\theta_2 = 0$, the extended Poisson family is obtained.

A natural question is which is the conjugated family of prior densities. Theorem 13 implies that this family is 3-parametric and the densities are

$$P_{prior}(\theta_1,\theta_2) =_B \exp\left(t_0 \ln C(e^{\theta_1},\theta_2) + t_1\theta_1 + t_2\theta_2\right) ,$$

where the parameters t_0, t_1 and t_2 have the following meaning: t_0 corresponds to the sample size; t_1 stands for the total sum of the observations, $\sum x_i$, and t_2 is the number of null observations, $\sum \delta(x_i)$. This family of priors contains both proper and improper priors because the t_i are arbitrary real numbers. The family of prior densities, as functions of the natural parameters of the family (12), i.e. (θ_1,θ_2), is in $B(\lambda_{\mathbb{R}^2})$. Finally, note that (12) may be expressed using the measure whose density is $(x!)^{-1}$ as a reference. In this case, the expression (12) remains the same but removing the factorial.

7. Conclusion

Classes of proportional σ-finite measures, including probability measures, have been structured as Bayes linear spaces. These classes can be represented by densities, including probability densities, likelihood functions and improper priors. The group operation, perturbation, is Bayes updating, thus defining a meaningful and interpretable structure. The affine subspaces are identified with extended exponential families, which include standard probability densities (or measures) and, additionally, infinite measures. Standard theorems of Bayesian statistics are revisited and slightly extended using this new algebraic-geometric point of view. The idea that Bayes theorem is the paradigm of information acquisition is now interpreted as an addition in the formal sense, being this possible because (proper and improper) probability densities and likelihood functions share the same Bayes space.

The presented framework permits a new interpretation of the standard probability theory, justifies the use of improper probability densities and opens up the study of some subspaces which may have richer structures with a metric or even a Hilbert space structure. The examples presented refer to quite usual probabilistic models, like normal and log-normal distributions; other distributions, although well-known and useful in practice (logistic normal, zero-inflated Poisson) need a more detailed mathematical development. The presented methodology, when applied to these examples, illustrates the new perspective introduced, namely how to deal with probability models in the framework of Bayes spaces. In particular, the idea that exponential families constitute an advanced mathematical tool in mathematical statistics, is here reduced to a very simple model, i.e. in the new framework they are linear affine subspaces.

Acknowledgements

This research has been supported by the Spanish Ministry of Education and Science under projects Ref.: MTM2009-13272 and Ref.: *Ingenio Mathematica (i-MATH)* No. CSD2006-00032 *(Consolider - Ingenio 2010)*, and by the *Agència de Gestió d'Ajuts Universitaris i de Recerca* of the *Generalitat de Catalunya* under project Ref: 2009SGR424.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 pp.

Aitchison, J., Barceló-Vidal, C., Egozcue, J. J. and Pawlowsky-Glahn, V. (2002). A concise guide for the algebraic-geometric structure of the simplex, the sample space for compositional data analysis. See Bayer *et al.*, pp. 387-392.

Ash, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York, NY (USA). 476 pp.

Barceló-Vidal, C., Martín-Fernández, J. A. and Pawlowsky-Glahn, V. (2001). Mathematical foundations of compositional data analysis. In G. Ross (Ed.), *Proceedings of IAMG'01 - The 6th annual conference of the Int. Ass. for Mathematical Geology*, pp. 20. CD-ROM.

Bauer, H. (1992). *Maß- und Integrationstheorie, 2 überarb. Auflage*. de Gruyter, Berlin (DE). 260 pp.

Bauer, H. (2002). *Wahrscheinlichkeitstheorie, 5 Auflage*. de Gruyter, Berlin (DE). 520 pp.

Bayer, U., Burger, H. and Skala, W. (Eds.) (2002). *Proceedings of IAMG'02 - The 8th annual conference of the Int. Ass. for Math. Geol.*, Volume I and II. Alfred-Wegener-Stiftung, Berlin (DE), ISSN 0946-8978, 1106 pp.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Pub. 378 pp.

Billheimer, D., Guttorp, P. and Fagan, W. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96(456), 1205-1214.

Birnbaum, A. (1962). On the foundations of statistical inference (with discussion). *Journal of the American Statistical Association*, 57, 269-326.

Eaton, M. L. (1983). *Multivariate Statistics. A Vector Space Approach*. John Wiley & Sons.

Egozcue, J. J., Díaz-Barrero, J. L. and Pawlowsky-Glahn, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. *Acta Mathematica Sinica*, 22(4), 1175-1182.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279-300.

Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 36, 210-271.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1-14.

Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods: an analysis for statisticians and interdisciplinary researchers*. Cambridge Series in Statistical and Probabilistical Mathematics. New York: Cambridge U. Press. 333 pp.

Martín-Fernández, J. A., Bren, M., Barceló-Vidal, C. and Pawlowsky-Glahn, V. (1999). A measure of difference for compositional data based on measures of divergence. In S. J. Lippard, A. Næss, and R. Sinding-Larsen (Eds.), *Proceedings of IAMG'99 - The 5th annual conference of the Int. Ass. for Math. Geol.*, Volume I and II, pp. 211-216. Tapir, Trondheim (N), 784 pp.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Barceló-Vidal, C. (2003). Distributions on the simplex. See Thió-Henestrosa and Martín-Fernández (2003).

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Martín-Fernández, J. A. (2002). Normal in $\Re^+$ vs lognormal in $\Re$. See Bayer *et al.* (2002), pp. 305-310.

Matusita, K. (1955). Decision rules based on the distance for problems of fit, two samples and estimation. *The Annals of Mathematical Statistics*, 26, 631-640.

Pawlowsky-Glahn, V. (2003). Statistical modelling on coordinates. See Thió-Henestrosa and Martín-Fernández (2003).

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5), 384-398.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2002). BLU estimators and compositional data. *Mathematical Geology*, 34(3), 259-274.

Robert, C. P. (2001). *The Bayesian Choice. A Decision Theoretic Motivation*. New York, NY (USA): Springer V. 436 pp.

Shao, J. (1999). *Mathematical Statistics*. Springer, New York (USA). 529 pp.

Small, C. G. and Leish, D. L. M. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. New York, NY (USA): Wiley-Interscience. 270 pp.

Thió-Henestrosa, S. and Martín-Fernández, J. A. (Eds.) (2003). *Compositional Data Analysis Workshop - CoDaWork'03, Proceedings*. Universitat de Girona, ISBN 84-8458-111-X, http://ima.udg.es/Activitats/CoDaWork03/.

Whaba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Soc. for Industrial & Applied Math. 165 pp.

Witting, H. (1985). *Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumpfang*. Stuttgart (DE): B. G. Teubner. 538 pp.

Appendix A. Proofs of theorems

Theorem 4

Proof. Perturbation: Since λ is σ-finite, there exists a family A_i, $i = 1,\ldots,n$, of sets increasing to Ω such that $\lambda(A_i) < +\infty$. Since μ and ν are in $\mathscr{M}(\lambda)$, they have $\lambda-$a.e. finite λ-equivalent densities f_μ and f_ν. Choose a version of these densities being everywhere finite and define families of sets $B_i := \{\omega \in \Omega : f_\mu(\omega) < i\}$, $C_i := \{\omega \in \Omega : f_\nu(\omega) < i\}$ increasing to Ω. Furthermore, consider the family of sets $D_i = A_i \cap B_i \cap C_i$; it is also increasing to Ω and

$$(\mu \oplus \nu)(D_i) =_B \int_{D_i} f_\mu f_\nu d\lambda \leq i^2 \lambda(A_i) < +\infty.$$

Thus, $\mu \oplus \nu$ is σ-finite.

Powering: Analogously, consider again the increasing family A_i, as well as the families $B_i := \{\omega \in \Omega : i^{-1} < f_\mu(\omega) < i\}$ and $C_i = A_i \cap B_i$. Then,

$$(\alpha \odot \mu)(A_i \cap B_i) = \int f_\mu^\alpha d\lambda \leq i^{|\alpha|} \lambda(A_i) < +\infty.$$

Thus, $\alpha \odot \mu$ is σ-finite. □

Theorem 5

Proof. According to the definition of Radon-Nikodym derivatives, expressions of $\oplus$ and $\odot$ using measures (1), (2), and using the respective densities, (6), (5), are equivalent. The operations are well defined on the equivalence classes since, for real constants k_1, k_2 and α,

$$(k_1 f_1 \oplus k_2 f_2)(x) = k_1 k_2 (f_1(x) f_2(x)) =_B (f_1 \oplus f_2)(x)\,,$$
$$(\alpha \odot k_1 f)(x) = k_1^\alpha f(x)^\alpha =_B (\alpha \odot f)(x)\,.$$

Linear space axioms follow from straightforward calculations:

- The neutral element is given by $\lambda =_B d\lambda/d\lambda =_B 1$.
- The opposite (negative) element is given by $(\ominus f_\mu) =_B (1/f_\mu) =_B d\lambda/d\mu$. □

Theorem 6

Proof. For measures, the equivalence relation $(=_B)$ does not depend on the reference measure λ; therefore, the quotient set $\mathscr{M}(\lambda)/(=_B)$ is equal to both $B(\mu)$ and $B(\lambda)$. In fact, any measure $\nu \in \mathscr{M}(\lambda)$ is represented in $B(\lambda)$ and $B(\mu)$ by B-equivalent densities; i.e. $\mu = k\lambda$, implies $d\lambda/d\mu = k$, $\lambda-$a.e., and then

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\lambda}\frac{d\lambda}{d\mu} = k\frac{d\nu}{d\lambda} \quad (\lambda\text{−a.e.})\,,$$

where $\lambda-$a.e. is equivalent to $\mu-$a.e. due to $\mu \in \mathscr{M}(\lambda)$. Therefore, operations $\oplus$ and $\odot$, expressed using densities, give proportional results when expressed in $B(\mu)$ or $B(\lambda)$. □

Theorem 7

Proof. Since $\mu \in \mathscr{M}(\lambda)$, $\mathscr{M}(\mu) = \mathscr{M}(\lambda)$. Furthermore, $(=_B)$-equivalence classes are the same in $\mathscr{M}(\mu)$ and in $\mathscr{M}(\lambda)$, and affine equivalence holds since there exists an affine mapping $g : B(\mu) \to B(\lambda)$, given by $g(\nu) :=_{B(\lambda)} \nu \ominus_\lambda \mu$, which is linear. Using the fact that $\ominus_\lambda \mu = d\lambda/d\mu$, and that any $\nu \in B(\mu)$ has the representation $(d\nu/d\mu)(d\mu/d\lambda)$ in $B(\lambda)$, linearity is given by:

$$g((\alpha \odot_\mu \nu_1) \oplus_\mu \nu_2) =_B g\left(\left(\frac{d\nu_1}{d\mu}\right)^\alpha \frac{d\nu_2}{d\mu}\right) =_B \underbrace{\left(\frac{d\nu_1}{d\mu}\frac{d\mu}{d\lambda}\frac{d\lambda}{d\mu}\right)^\alpha}_{g(\nu_1)} \underbrace{\frac{d\nu_2}{d\mu}\frac{d\mu}{d\lambda}\frac{d\lambda}{d\mu}}_{g(\nu_2)}$$
$$=_B (\alpha \odot_\lambda g(\nu_1)) \oplus_\lambda g(\nu_2)\,,$$

where the subscripts of $\oplus$ and $\odot$ indicate the reference measure of the space where the operation is carried out. □

Theorem 8

Proof.

1. $\mu =_B \nu$ is equivalent to $\mu(\Omega) = k\nu(\Omega)$; therefore, μ, ν are either finite or infinite and then B_P and B_I are well defined and they constitute the whole space.
2. For any densities f, g, in $B_P(\lambda)$ and for any value $0 \leq \alpha \leq 1$, the statement is equivalent to
$$(\alpha \odot f) \oplus ((1-\alpha) \odot g) = \int f^\alpha g^{1-\alpha} d\lambda \leq \int f d\lambda + \int g d\lambda < +\infty \,.$$
3. Boundedness is preserved by arbitrary powering and perturbation with bounded values.
4. The same holds for upper boundedness as long as the exponents are positive.
5. It follows from the inequality $fg < bf$ ($\lambda-$a.e.).
6. It follows from the inequality $f/g < f/b$ ($\lambda-$a.e.).
7. ($\Rightarrow$): $\nu(\Omega)$ does not depend on λ.
($\Leftarrow$): λ and μ are λ-equivalent and then $\mu \in B(\lambda)$.
8. If $\mu \in B_b(\lambda)$, then $b_1^{-1} \leq d\mu/d\lambda < b_1$, and if $\nu \in B_b(\mu)$, then $b_2^{-1} \leq d\nu/d\mu < b_2$; combining both expressions, $(b_1 b_2)^{-1} \leq d\nu/d\lambda = (d\nu/d\mu)(d\mu/d\lambda) \leq b_1 b_2$ and then $\nu \in B(\lambda)$.
9. ($\Rightarrow$): If $\nu \in B_b(\mu)$ with density f, $0 < b^{-1} \leq f \leq b$ and $\int f \, d\mu \leq b\mu(\Omega) < +\infty$.
($\Leftarrow$): $\nu \in B_b(\mu) \subset B(\mu)$ implies $+\infty > \int f d\mu \geq b^{-1}\mu(\Omega)$, then $\mu(\Omega) < +\infty$.
10. Similar to the previous statement. □

Theorem 9

Proof. Let $\mu_{\vec{\boldsymbol{\alpha}}} \in \mathrm{Exp}_B(\lambda, g, \vec{T}, \vec{\boldsymbol{\theta}})$ be a measure. By definition $\mu_{\vec{\boldsymbol{\alpha}}}$ is λ-equivalent and $\mu_{\vec{\boldsymbol{\alpha}}} \in B(\lambda)$. Then, it can be expressed as

$$\mu_{\vec{\boldsymbol{\alpha}}} =_B g \oplus \bigoplus_{j=1}^{k} (\theta_j(\vec{\boldsymbol{\alpha}}) \odot V_j(x)) \,,$$

with $V_j =_B \exp(T_j)$. Therefore, the exponential family corresponds to the affine subspace of $B(\lambda)$

$$g \oplus \mathrm{span}\{V_j, j = 1, \ldots, k\} \,,$$

where the natural parameters $\theta_j(\vec{\boldsymbol{\alpha}})$ are the coordinates of $\mu_{\vec{\boldsymbol{\alpha}}}$ with respect to the basis elements V_j. □

Theorem 10

Proof. Let $g \in S$ be a density and V_j, $j = 1, 2, \ldots, k$, be a basis of the subspace $S \ominus g$. Any element $\mu \in S$ is expressed as $\mu =_B g \oplus \bigoplus_{j=1}^{k} (\alpha_j \odot V_j)$, thus spanning exactly S. Then, $\mu \in \mathrm{Exp}_B(\lambda, g, \ln\vec{V}, \vec{Id})$, with $\ln\vec{V} = (\ln V_1, \ldots, \ln V_k)$ and $\vec{Id}$ the identity mapping. The parametrisation is strict, since the coordinates with respect to a basis are unique. □

Theorem 11

Proof. The statement is proven if, L_{x_i} is a τ-equivalent density of a σ-finite and τ-equivalent measure $P_\theta(x_i)$-a.e. For $\theta \in \Theta$, $L_{x_i} > 0$ since $P_\theta \in B(\lambda)$. Thus, it is τ-equivalent. It is in $B(\tau)$ if it corresponds to a σ-finite measure. To prove that L_{x_i} is a density of a σ-finite measure, consider any finite measure $\tau' \in B_P(\tau)$. If $P(x_i, \theta)$ is the joint probability distribution of X_i and θ constructed from τ' as marginal distribution, then

$$L_{x_i}(\theta) = \frac{dP(x_i, \theta)}{d\tau'(\theta) d\lambda(x_i)},$$

because P_θ is the conditional distribution and τ' plays the role of a marginal distribution for θ. Fubini theorem implies $\int L_{x_i} d\tau < +\infty$ (λ-a.e.), or, equivalently, P_θ-a.e. Then, $L_{x_i} \in B(\tau')$ and represents a finite measure μ_{x_i} (P_θ-a.e.). According to Theorem 7 on shift of origin, from $B(\tau)$ to $B(\tau')$, we get $L_{x_i} =_{B(\tau)} \mu_{x_i} \ominus_\tau \tau'$ and thus $L_{x_i} \in B_P(\tau)$. □

Theorem 13

Proof. The likelihood function can be written

$$\begin{aligned} L_{\vec{x}}(\vec{\boldsymbol{\theta}}) &= C^n(\vec{\boldsymbol{\theta}}) \cdot \prod_{i=1}^{n} g(x_i) \cdot \exp\left(\sum_{i=1}^{n} \sum_{j=1}^{k} \theta_j T_j(x_i) \right) \\ &=_{B(\tau)} C^n(\vec{\boldsymbol{\theta}}) \cdot \exp\left(\sum_{j=1}^{k} \theta_j \left[\sum_{i=1}^{n} T_j(x_i) \right] \right). \end{aligned} \tag{13}$$

If $C(\vec{\boldsymbol{\theta}})$ is in the span of $\exp(\vec{\boldsymbol{\theta}})$, $L_{\vec{x}}(\vec{\boldsymbol{\theta}})$ corresponds to a k-dimensional subspace of $B(\tau)$ with $g^*(\vec{\boldsymbol{\theta}}) =_{B(\tau)} 1$, $\vec{\boldsymbol{\theta}}^* = \vec{\boldsymbol{\theta}}$ and $\vec{T}^* = \vec{T}$. Otherwise, taking $g^*(\vec{\boldsymbol{\theta}}) = 1$, $\vec{\boldsymbol{\theta}}^* = (\ln C(\vec{\boldsymbol{\theta}}), \vec{\boldsymbol{\theta}})$, and $\vec{T}^*(\vec{x}) = (n, \sum_{i=1}^{n} \vec{T}(x_i))$, Eq. $L_{\vec{x}}(\vec{\boldsymbol{\theta}})$ corresponds to a $(k+1)$-dimensional subspace of $B(\tau)$. In both cases, Theorem 9 implies the statement. □

Theorem 14

Proof. The family $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$ is a subspace of $B(\tau)$ because $g^* =_B 1$. Since subspaces are invariant under perturbation of elements of the subspace, the posterior $P_{post}(\vec{\boldsymbol{\theta}})$ is in the subspace. □

Theorem 15

Proof. The likelihood $L_{\vec{x}}(\vec{\boldsymbol{\theta}})$, as a function of $\vec{\boldsymbol{\theta}}$, is in the extended exponential family $\mathrm{Exp}_{B(\tau)}(\tau, g^*, \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$ that has been identified as a subspace of $B(\tau)$. Application of Bayes theorem is a perturbation, i.e. a shifting, and the result is the affine space $\mathrm{Exp}_{B(\tau)}(\tau, P_{prior}(\vec{\boldsymbol{\theta}}), \vec{T}^*, \vec{\boldsymbol{\theta}}^*)$, where the origin coincides with P_{prior} because $g^* =_{B(\tau)} 1$. □

Theorem 16

Proof. The posterior density in the extended exponential family is expressed as

$$P_{post}(\vec{\boldsymbol{\theta}}) =_{B(\tau)} h \oplus \exp\left(\sum_{j=1}^{k} S_j(\vec{x})\theta_j\right) .$$

Combining this expression with the Bayes formula, the likelihood function is

$$L_{\vec{x}}(\vec{\boldsymbol{\theta}}) =_{B(\tau)} (h \ominus P_{prior}(\vec{\boldsymbol{\theta}})) \oplus \exp\left(\sum_{j=1}^{k} S_j(\vec{x})\theta_j\right) .$$

In $B(\lambda)$ it can be rewritten as

$$L_{\vec{x}}(\vec{\boldsymbol{\theta}}) =_{B(\lambda)} \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{k} T_j(x_i)\theta_j\right) ,$$

where $S_j(\vec{x}) = \sum_{i=1}^{n} T_j(x_i)$. The existence of the statistics T_j comes from the multiplicative form of the likelihood function and the fact that the expression should be valid for any arbitrary n. Therefore,

$$L_x(\vec{\boldsymbol{\theta}}) =_{B(\lambda)} 1 \cdot \exp\left(\sum_{j=1}^{k} T_j(x)\theta_j\right) ,$$

where the perturbation of k terms may collapse in $k_1 \leq k$ terms for equal T_j's. □

SORT 34 (2) July-December 2010, 223-238

Optimal inverse Beta(3,3) transformation in kernel density estimation

Catalina Bolancé*
University of Barcelona

Abstract

A double transformation kernel density estimator that is suitable for heavy-tailed distributions is presented. Using a double transformation, an asymptotically optimal bandwidth parameter can be calculated when minimizing the expression of the asymptotic mean integrated squared error of the transformed variable. Simulation results are presented showing that this approach performs better than existing alternatives. An application to insurance claim cost data is included.

MSC: 62G07, 62G32

Keywords: Kernel density estimation, transformations, Beta density, right skewness.

1. Introduction

Kernel density estimation is nowadays a classical approach to study the form of a density with no assumption on its global functional form.

Let $X_1,\ldots,X_n$ a random sample of *iid* observations of a random variable with density function f, then the kernel density estimator at point x is:

$$\hat{f}_c(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(x - X_i), \tag{1}$$

where b is the bandwidth or smoothing parameter, $K_b(t) = \frac{1}{b}K\left(\frac{t}{b}\right)$ and K is the kernel function, usually it is a symmetric density function bounded or asymptotically bounded

* Dept. of Econometrics. University of Barcelona. Diagonal, 690 08034 Barcelona, Spain.
Email: bolance@eco.ub.es. Support from grant SEJ2007-63298 is acknowledged.
Received: June 2010
Accepted: October 2010

and centred at zero. In this work I use the Epanechnikov kernel, Silverman (1986) proves that this kernel is optimal for kernel density estimator. The Epanechnikov kernel is:

$$k(t) = \begin{cases} 0.75\left(1-t^2\right) & \text{si } |t| \leq 1 \\ 0 & \text{si } |t| > 1 \end{cases}$$

Silverman (1986) or Wand and Jones (1995) provide an extensive review of classical kernel estimation. In order to implement kernel density estimation both K and b need to be chosen. The optimal choice for the value of b depends inversely on the sample size, so the larger the sample size, the smaller the smoothing parameter and conversely.

When the shape of the density to be estimated is symmetric and has a kurtosis that is similar to the kurtosis of the normal distribution, then it is possible to calculate a smoothing parameter b that provides optimal smoothness or is close to optimal smoothness over the whole domain of the distribution. However, when the density is asymmetric, it is not possible to calculate a value for the smoothing parameter which captures both the mode of the density shape and the tail behaviour. In fact, optimal smoothness in the tail is much larger than in the main mode and this is due to the fact that available sampling information in the mode is much more abundant than in the tail of the density, where there are not many observations.

The majority of economic variables that measure expenditures or costs have a strong asymmetric behaviour to the right, so that classical kernel density estimation is not efficient in order to estimate the values of the density in the right tail part of the density domain. This is due to the fact that the smoothing parameter which has been calculated for the whole domain function is too small for the density in the tail. Using a variable bandwidth can be a convenient solution, but this approach has many difficulties as discussed by Jones (1990). Our aim is to propose a double transformation kernel density estimator, where the bandwidth is optimal and can be chosen automatically. The optimal bandwidth has a straightforward expression and it is obtained by minimizing the asymptotic mean integrated squared error.

An alternative to kernel estimation defined in (1) is transformation kernel estimation that is based on transforming the data so that the density of the transformed variable has a symmetric shape, so that it can easily be estimated using a classical kernel estimation approach. We say it can be easily estimated in the sense that using a Gaussian kernel or an Epanechnikov kernel, an optimal estimate of the smoothing parameter can be obtained by minimizing an error measure over the whole density domain. In the specialized literature several transformation kernel estimators have been proposed, and their main difference is the type of transformation family that they use. For instance, Wand *et al.* (1991), Bolancé *et al.* (2003), Clements *et al.* (2003) and Buch-Larsen *et al.* (2005) propose different parametric transformation families that they all make the transformed distribution more symmetric that the original one, which in many applications has usually a strong right-hand asymmetry. Also Bolancé *et al.* (2008) used

the transformation kernel estimation to approximate the conditional tail expectation risk measure.

Given a density estimator $\widehat{f}$ of a density f, the Mean Integrated Squared Error (MISE) is defined as:

$$MISE\left(\hat{f}\right) = E\left(\int_{-\infty}^{+\infty}\left(\widehat{f}(t) - f(t)\right)^2 dt\right).$$

Let $T(\cdot)$ a concave transformation, the transformed sample is $Y_1 = T(X_1), \ldots, Y_n = T(X_n)$, the classical kernel estimator of the transformed variable is:

$$\hat{f}_c(y) = \frac{1}{n}\sum_{i=1}^{n} K_b(y - Y_i) = \frac{1}{n}\sum_{i=1}^{n} K_b(T(x) - T(X_i)) \tag{2}$$

and the transformation kernel estimator of the original variable is:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(T(x) - T(X_i))\, T'(x). \tag{3}$$

Wand *et al.* (1991) show that there exists a relationship between the value of *MISE* obtained for the classical kernel estimator of the transformed variable and the *MISE* obtained with the transformation kernel estimator of the original variable. They also show that there exists an optimal transformation that minimizes both expressions.

Based on the work by Buch-Larsen *et al.* (2005), Bolancé *et al.* (2008) proposed a double transformation with the purpose of obtaining a transformed variable whose density is as close as possible to a density that maximizes smoothness $\int \{f''(x)\}^2 dx$ and at the same time that minimizes the asymptotic Mean Integrated Squared Error ($A - MISE$) of the kernel estimator defined in (1) and obtained with the transformed observations. Terrell and Scott (1985) showed that among the vast family of densities with domain D that have a Beta distribution, one of them has the largest possible smoothness.

Since the density of a Beta distribution in the bounds of its domain is zero, the bias of kernel estimation near the boundaries of the domain is strictly positive, and therefore this implies a larger bias in the transformation kernel estimation in the extremes of the density of the original variable (in the right tail and in the values near the minimum). In order to correct for this positive bias, Bolancé *et al.* (2008) proposed to transform their data into a new set of data so that they have a density that is similar to the Beta density in a domain in the interior of D. Then they correct the resulting density estimate so that it integrates to one, but in their contribution they do not indicate how to optimize this second transformation. In the next section, a method based on minimizing $A - MISE$ is proposed. One of its main features is that it can become fully automated, which is very suitable for practical applications.

Let $g(\cdot)$ and $G(\cdot)$ be the density and distribution functions of a Beta random variable, which we denote by $B(\beta,\beta)$ with domain in $[-\alpha,\alpha]$, if Z is a random variable with a uniform distribution, then $Y = G^{-1}(Z)$ is a random variable with distribution $B(\beta,\beta)$. The method proposed by Bolancé *et al.* (2008) suggests to do a first transformation on the original sample of observations $X_1,\ldots,X_n$ so that $Z_i = T(X_i)$, $i = 1,\ldots,n$. If $T(\cdot)$ is a cumulative distribution function then Z_i, $i = 1,\ldots,n$ can be a sample of independent observations that are close to have been generated by a uniform distribution. Then they define l as a probabililty close to 1, namely 0.98 or 0.99, so that $\tilde{T}(X_i) = \tilde{Z}_i = (2l-1)Z_i + (1-l)$ and, therefore, the density that is associated with the data generating process $Y_i = G^{-1}\left(\tilde{Z}_i\right)$ coincides with the density function of a Beta density, $B(\beta,\beta)$ in a domain $[-a,a]$, where $\alpha > a = G^{-1}(l)$. Then, the resulting transformation kernel estimator, where $'$ denotes the first derivative, is:

$$\begin{aligned}\widehat{f}(x) &= \frac{1}{(2l-1)n}\sum_{i=1}^{n} K_b\left(G^{-1}\left(\tilde{T}(x)\right) - G^{-1}\left(\tilde{T}(X_i)\right)\right)\left(G^{-1}\right)'\left(\tilde{T}(x)\right)\tilde{T}'(x)\\ &= \frac{1}{n}\sum_{i=1}^{n} K_b\left(G^{-1}\left(\tilde{T}(x)\right) - G^{-1}\left(\tilde{T}(X_i)\right)\right)\left(G^{-1}\right)'\left(\tilde{T}(x)\right)T'(x). \qquad (4)\end{aligned}$$

We note that the optimality of (4) depends on whether the first transformation $T(\cdot)$ is successfully transforming the data into a sample that is likely to have been generated by a Uniform$(0,1)$. It is obvious that the transformation $T(\cdot)$ must be a distribution function. Bolancé *et al.* (2008) propose to use the generalized Champernowne cdf:

$$T_{\alpha,M,c}(x) = \frac{(x+c)^{\alpha} - c^{\alpha}}{(x+c)^{\alpha} + (M+c)^{\alpha} - 2c^{\alpha}} \qquad x \geq 0, \qquad (5)$$

with parameters $\alpha > 0$, $M > 0$ and $c \geq 0$, that can be estimated by maximum likelihood. This is certainly a flexible distribution, because it can have many shapes near zero and also different behaviours in the tail. Degen and Embrechts (2008) analyzed the tail modified Champernowne distribution convergence to the tail behaviour supposed by extreme value theory, and they concluded that convergence is stronger if we compare it to the tail distribution for the Loggamma, the g-and-h and the Burr and lighter if we compare it to the Generalized Beta distribution (GB2).

In this work we propose a method to find an asymptotically optimal value for l, that is obtained when one finds the Beta truncated distribution with density $\frac{g(\cdot)}{(2l-1)}$, defined on $[-a,a]$, with $a = G^{-1}(l)$, whose kernel estimation minimizes *MISE* asymptotically. This result is developed in Section 2. Section 3 presents the results of a simulation study that uses the same samples as in Buch-Larsen *et al.* (2005) and in Bolancé *et al.* (2008). By means of the results of the simulation we analyze the behaviour of the estimation method that is being proposed and we see that the value of the optimal choice for l considerably reduces the distance between the true theoretical density and the density

estimate for all the asymmetric shapes that have been analyzed and, in many cases, also if the sample size is small. In Section 4 we show an application to data on costs arising from automobile insurance claims. These data were also used by Bolancé *et al.* (2009). Finally, in Section 5 we conclude.

2. Asymptotically optimal truncated inverse Beta transformation

Terrell and Scott (1985, Lemma 1) showed that $B(3,3)$ defined on the domain $(-1/2,1/2)$ has $\int \{g''(t)\}^2 dt$ minimal within the set of Beta densities with same support, where $g(\cdot)$ is the pdf and is given by:

$$g(t) = \frac{15}{8}\left(1-4t^2\right)^2, -\frac{1}{2} \leq t \leq \frac{1}{2} \tag{6}$$

and $G(\cdot)$ is the cdf and is given by:

$$G(t) = \frac{1}{8}\left(4-9t+6t^2\right)(1+2t)^3. \tag{7}$$

Using the Epanechnikov kernel for the upper bound (or the lower bound since the domain of the distribution $B(3,3)$ is symmetric) the expectation of the classical kernel estimation is (see, Wand and Jones 1995, p. 47):

$$\begin{aligned}\int_{-1}^{0} K(t)\, g\left(\frac{1}{2}-bt\right) dt &= \int_{-1}^{0} \frac{3}{4}\left(1-(t)^2\right)\frac{15}{8}\left(1-4\left(\frac{1}{2}-bt\right)^2\right)^2 dt \\ &= 1.2857b^4 + 3.75b^3 + 3b^2 > 0 \text{ if } b > 0. \end{aligned} \tag{8}$$

The value of the density defined in (6) in the boundaries of the domain is zero, however, as we have noted in (8), the value of the classical kernel estimation of the density is positive $\forall b > 0$, and therefore $\hat{f}_c(x)$ over-estimates the beta density in the tails.

Silverman (1986) shows that asymptotically the MISE for (1) is:

$$A-MISE\left\{\hat{f}_c\right\} = \frac{1}{4}b^4k_2^2\int f''(x)^2\,dx + \frac{1}{nb}\int K(t)^2\,dt,$$

where $k_2 = \int t^2K(t)\,dt$. The asymptotically optimal bandwidth is:

$$b^{opt} = \left(\frac{\int K(t)^2\,dt}{k_2^2\int f''(x)^2\,dx}\right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

replacing b^{opt} in $A-MISE\left\{\hat{f}_c\right\}$ we obtain the value of $A-MISE$ for the asymptotically optimal bandwidth:

$$A-MISE^{*}\left(\hat{f}_c\right)=\frac{5}{4}k_2^{\frac{2}{5}}\left(\int K(t)^2\,dt\right)^{\frac{4}{5}}\left(\int f''(x)^2\,dx\right)^{\frac{1}{5}}n^{-\frac{4}{5}}. \tag{9}$$

Let Y be a transformed random variable with distribution $B(3,3)$. Let $\frac{g(y)}{2l_a-1}$, with $l_a=G(a)$, be the truncated Beta density in the domain $[-a,a]$. If one just uses the same development that is being used to obtain (9), a value for $A-MISE^{*}\left(\hat{f}_c(x),a\right)$ can easily be obtained. Replacing in Silverman's $A-MISE$ proof $g(y)$ by $\frac{g(y)}{2l_a-1}$ we obtain:

$$A-MISE\left\{\hat{g}_c,a\right\}=\frac{1}{4}b^4\frac{k_2^2}{(2l_a-1)^2}\int_{-a}^{+a}g''(x)^2\,dx+\frac{1}{nb}\int K(t)^2\,dt,$$

then

$$b^{opt}(a)=\left(\frac{\int K(t)^2\,dt}{\frac{k_2^2}{(2l_a-1)^2}\int_{-a}^{+a}g''(x)^2\,dx}\right)^{\frac{1}{5}}n^{-\frac{1}{5}}, \tag{10}$$

and replacing $b^{opt}(a)$ in $A-MISE\left\{\hat{f}_c,a\right\}$ we obtain:

$$A-MISE^{*}\left\{\hat{g}_c,a\right\}=\frac{5}{4}k_2^{\frac{2}{5}}\left(\int K(t)^2\,dt\right)^{\frac{4}{5}}(2l_a-1)^{-\frac{2}{5}}\left(\int_{-a}^{+a}g''(x)^2\,dx\right)^{\frac{1}{5}}n^{-\frac{4}{5}}.$$

We then analyze the behaviour of $A-MISE^{*}\left\{\hat{g}_c,a\right\}$ as a function of a in order to estimate the truncated density $\frac{g(y)}{2l_a-1}$ whenever the objective is that the distribution of the transformed variable is $B(3,3)$. Using the Epanechnikov's kernel $K(t)=\frac{3}{4}\left(1-t^2\right)$, $|t|\leq 1$ for the density of a $B(3,3)$ we obtain:

$$A-MISE^{*}\left\{\hat{g}_c,a\right\}=\frac{5}{4}\left(\frac{9}{125}\right)^{\frac{2}{5}}\left(\frac{360a\left(-40a^2+144a^4+5\right)}{\left(\frac{1}{4}a\left(-40a^2+48a^4+15\right)\right)^2}\right)^{\frac{1}{5}}n^{-\frac{4}{5}}. \tag{11}$$

If we also analyze the shape of expression (11), we observe that there exists a value of a that minimizes the corresponding expression for $A-MISE^{*}$. In Figure 1 we show a plot of (11) as a function of a, where we have eliminated the effect of the sample size factor ($n^{-\frac{4}{5}}$).

As a result, there exists a truncated density $\frac{g(y)}{2l_{a^*}-1}$ that depends on an optimal a which is related to $B(3,3)$ that minimizes (11). The objective of our proposed transformation kernel estimation method is to obtain a sample of transformed observations whose

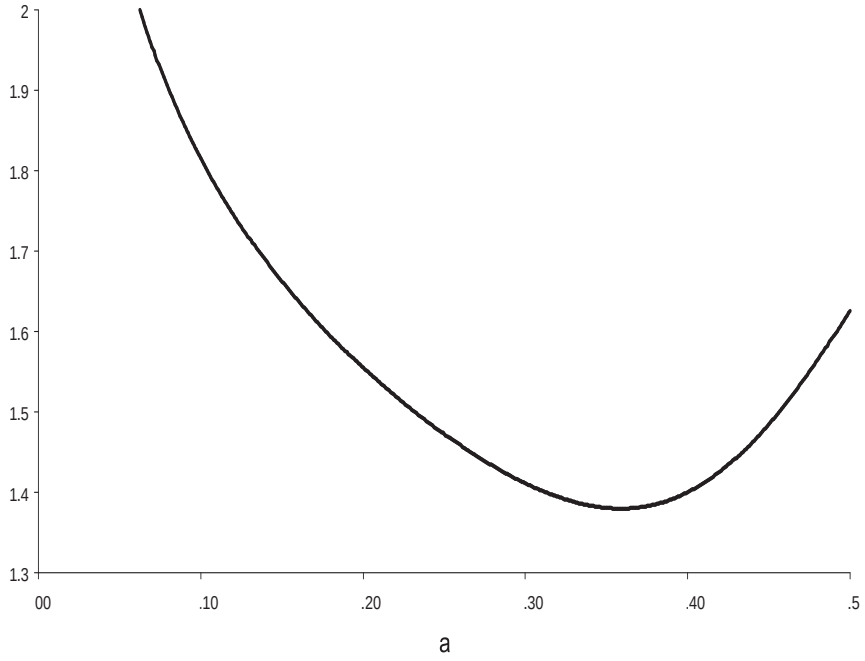


Figure 1: $A-MISE^{*}_{B(3,3)}\{\hat{g}_c,a\}n^{\frac{4}{5}}$ *vs a.*

density is as close as possible to an optimally truncated Beta density, so that the optimality of the kernel estimation of the transformed variable is transferred to an optimal transformation kernel estimation of the original variable. Then we propose:

$$\widehat{f}^{*}(x)=\frac{1}{n}\frac{\sum_{i=1}^{n}K_b\left(G^{-1}\left(\tilde{T}^{*}(x)\right)-G^{-1}\left(\tilde{T}^{*}(X_i)\right)\right)\left(G^{-1}\right)'\left(\tilde{T}^{*}(x)\right)\tilde{T}^{*\prime}(x)}{(2l_{a^*}-1)}$$

$$=\frac{1}{n}\sum_{i=1}^{n}K_b\left(G^{-1}\left(\tilde{T}^{*}(x)\right)-G^{-1}\left(\tilde{T}^{*}(X_i)\right)\right)\left(G^{-1}\right)'\left(\tilde{T}^{*}(x)\right)T'(x) \qquad (12)$$

where $\tilde{T}^{*}(X_i)=\tilde{Z}_i^{*}=(2l_{a^*}-1)Z_i+(1-l_{a^*})$. Holding n fixed, when we minimize (11) we obtain an optimal a, which we call a^{*} equal to 0.389121. Therefore, $l_{a^*}=G(0.389121)=0.98854$. We call the estimator defined in (12) optimal double transformation kernel density estimator or optimal Kernel Inverse Beta Modified Champernowne Estimator (KIBMCE) if we use the same name given in Bolancé *et al.* (2008).

In order to obtain the estimator in (12) the procedure is:

1. With the sample of observations $X_1,\ldots,X_n$ we estimate parameters α, M and c of the generalized Champernowne by maximum likelihood (see, for instance, Burch-Larsen *et al.*, 2005) and calculate (5) $Z_i=T_{\hat{\alpha},\hat{M},\hat{c}}(X_i)$ and $\tilde{T}^{*}(X_i)=\tilde{Z}_i^{*}=(2\cdot 0.98854-1)Z_i+(1-0.98854)$.

2. Calculate $Y_i = G^{-1}\left(\tilde{T}^*(X_i)\right)$ and obtain the classical kernel estimator $\hat{f}_c(y)$ defined in (1). The smoothing parameter b^* is estimated by the value that is asymptotically optimal when estimating a $B(3,3)$ on the domain $(-a^*, a^*)$, and therefore its expression is:

$$b^* = k_2^{-\frac{2}{5}} \left(\int_{-1}^{1} K(t)^2 \, dt \int_{-a^*}^{a^*} g(y) \, dy \right)^{\frac{1}{5}} \left(\int_{-a^*}^{a^*} \left\{ g''(y) \right\}^2 dy \right)^{-\frac{1}{5}} n^{-\frac{1}{5}}$$

$$= 0.5416079 n^{-\frac{1}{5}}. \tag{13}$$

The difference between the smoothing parameter $b^{opt}(a)$ in (10) and b^* in (13) is that first is optimal for the classical kernel estimation of truncate Beta density and second is optimal for the classical kernel estimation of Beta density in $[-a^*, a^*]$.

3. Obtain the optimal double transformation kernel estimator in (12) as:

$$\widehat{f^*}(x) = \hat{f}_c(y) \left(G^{-1}\right)' \left(\tilde{T}^*(x)\right) T'(x).$$

It is obvious that the estimator in (12) is optimal if the transformed random variable $Z = T(X)$ is distributed as a Uniform$(0,1)$, and this certainly depends on the quality of the generalized Champernowne cdf defined in (5) and how well it approximates the original variable. This is going to be discussed in the next section, where simulation results are also shown.

Next we are going to present a simulation study where we show to what extend, for finite sample, and with the transformation kernel estimation expressed in (12) the results shown in Buch-Larsen *et al.* (2005) can be improved. Therefore it also improves Wand *et al.* (1991) and Clements *et al.* (2003).

3. Simulation study

This section presents a comparison of our inverse beta double transformation method with the results presented by Buch-Larsen *et al.* (2005) based only on the modified Champernowne distribution. Our objective is to show that the second transformation, that is based on the inverse of a Beta optimal truncated distribution, improves density estimation for a wide range of asymmetric densities that are commonly found in practice.

In this work we analyze the same simulated samples as in Buch-Larsen *et al.* (2005) and Bolancé *et al.* (2008), which were drawn from four distributions with different tails and different shapes near 0. The distributions and the chosen parameters are listed in Table 1.

Table 1: *Distributions in simulation study.*

Distribution	Density	Parameters
Mixture of p Lognormal (μ,σ) and $(1-p)$ Pareto (λ,ρ,c)	$f(x) = p\frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} + + (1-p)(x-c)^{-(\rho+1)}\rho\lambda^{\rho}$	$(p,\mu,\sigma,\lambda,\rho,c)$ $=(0.7,0,1,1,1,-1)$ $=(0.3,0,1,1,1,-1)$ $=(0.1,0,1,1,1,-1)$ $=(0.9,2.5,0.5,1,1,-1)$
Lognormal (μ,σ)	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$	$(\mu,\sigma)=(0,0.5)$
Weibull (γ)	$f(x)=\gamma x^{(\gamma-1)}e^{-x^{\gamma}}$	$\gamma=1.5$
Truncated logistic	$f(x)=\frac{2}{s}e^{\frac{x}{s}}\left(1+e^{\frac{x}{s}}\right)^{-2}$	$s=1$

In Figure 2 we present the result of the ratio between the distribution function $F(x)$ that is associated to each of the densities in Table 1 and the Champernowne distribution $T_{\hat{\alpha},\hat{M},\hat{c}}(x)$ that is estimated by means of a sample with size 1000, obtained from each of the five distribution. The right-hand plots focus on the ratio in the tail. In Table 2, we show the distance measures L_1 and L_2 between $F(x)$ and $T_{\hat{\alpha},\hat{M},\hat{c}}(x)$:

$$L_1\left(F,T_{\hat{\alpha},\hat{M},\hat{c}}\right)=\int_{-\infty}^{+\infty}\left|T_{\hat{\alpha},\hat{M},\hat{c}}(t)-F(t)\right|dt$$

and

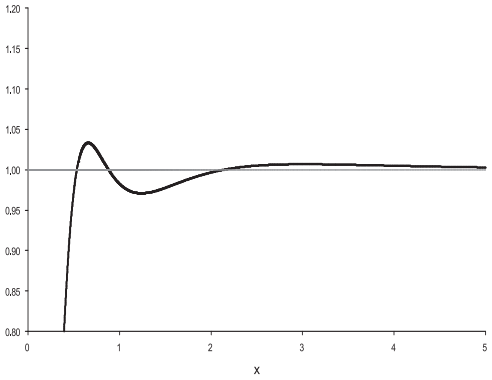
$$L_2\left(F,T_{\hat{\alpha},\hat{M},\hat{c}}\right)=\int_{-\infty}^{+\infty}\left(T_{\hat{\alpha},\hat{M},\hat{c}}(t)-F(t)\right)^2dt.$$

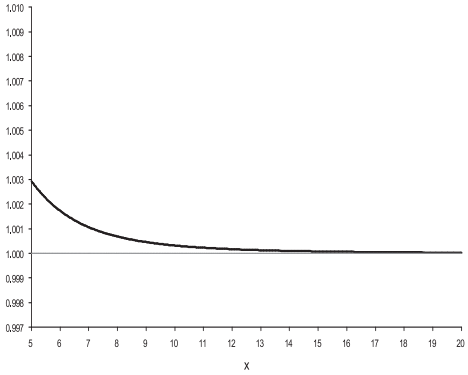
Table 2: *Distance between the true distribution and the Champernowne distribution.*

	Lognormal	Log-Pareto		Weibull	Tr. Logist.
		$p=0.7$	$p=0.3$		
L_1	0.0445	1.4423	2.1270	0.0422	0.0940
L_2	0.0225	0.0409	0.0544	0.0240	0.0343

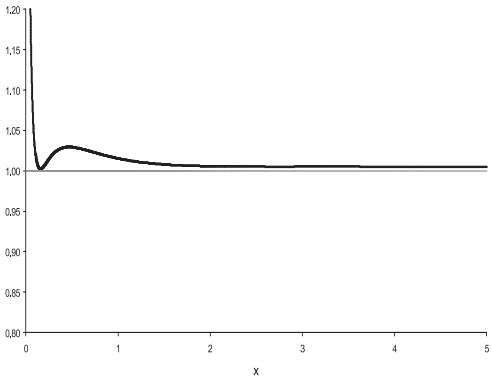
It is obvious that the improvement in the KIBMCE method with respect to the Kernel Modified Champernowne Estimator (KMCE) proposed by Buch-Larsen *et al.* (2005) is larger in those cases where the shape of the true cdf is similar to the Champernowne. In

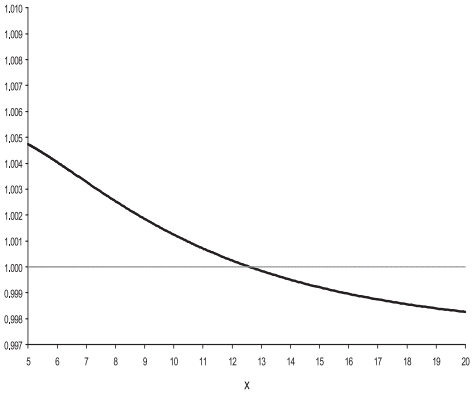
a) Lognormal



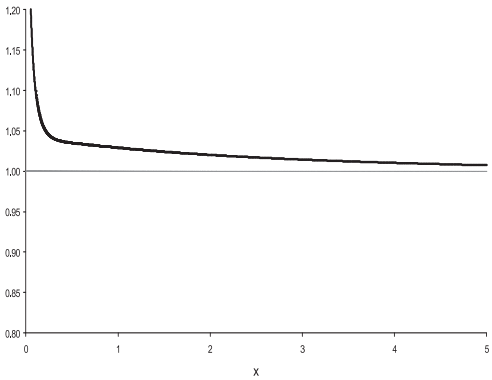


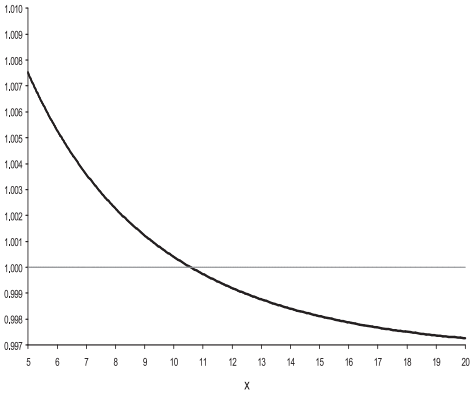
b) 70% Lognormal-30% Pareto



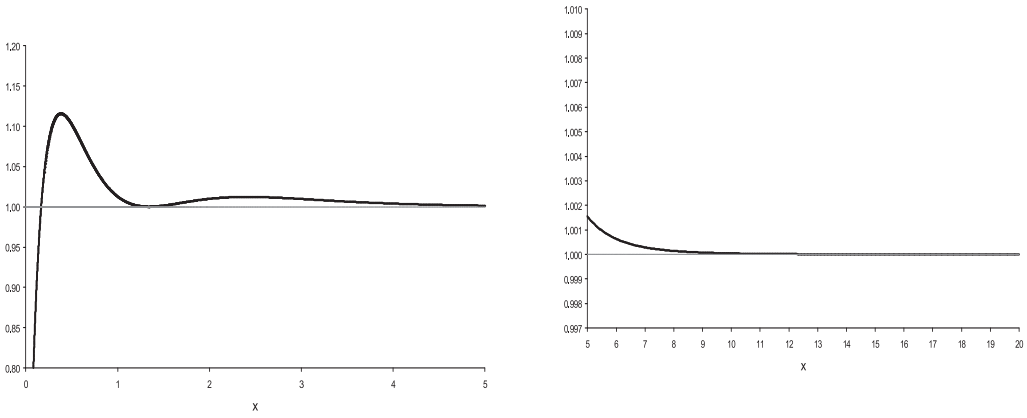


c) 30% Lognormal-70% Pareto





d) Weibull



e) Truncated Logistic

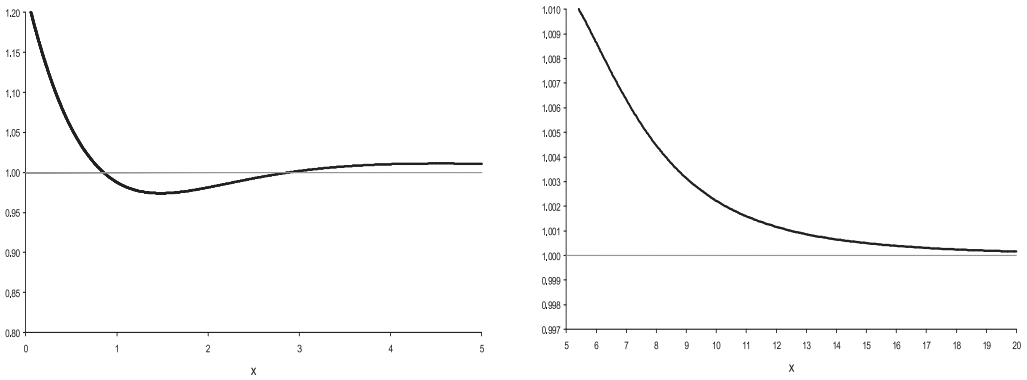


Figure 2: *Ratio of $F(x)$ and $T_{\hat{\alpha},\hat{M},\hat{c}}(x)$ in the $(0,5)$ domain interval on the left and in the $(5,20)$ domain interval on the right, for five distributions given in Table 1.*

the case of a mixture between a lognormal and a Pareto, Figures 2b and 2c show that the Champernowne distribution tends more rapidly to one that the true cdf. and this can also be seen when looking at the values of the L_1 distance between the two functions. The results of Figure 1 and Table 2 show us that the improvement in KIBMCE is larger in the estimation of a density that has a Lognormal, a Weibull and a Truncated Logistic shape.

Buch-Larsen *et al.* (2005) evaluate the performance of the KMCE estimators compared to the estimator described by Clements *et al.* (2003) the estimator described by Wand *et al.* (1991) and the estimator described by Bolancé *et al.* (2003). The Champernowne transformation substantially improve the results from previous authors. Bolancé *et al.* (2008, 2009) compare his truncated inverse beta second transformation with

Table 3: *The estimated error measures for KMCE and KIBMCE.*

				Lognormal	Log-Pareto		Weibull	Tr. Logist.
			l_{a^*}		$p=0.7$	$p=0.3$		
N =100	L_1	KIBMCE	0.9885	0.1348	0.1240	0.1202	0.1391	0.1246
		KMCE		0.1363	0.1287	0.1236	0.1393	0.1294
	L_2	KIBMCE	0.9885	0.1001	0.0851	0.0853	0.1095	0.0739
		KMCE		0.1047	0.0837	0.0837	0.1084	0.0786
	WISE	KIBMCE	0.9885	0.0992	0.0819	0.0896	0.0871	0.0969
		KMCE		0.1047	0.0859	0.0958	0.0886	0.0977
N =1000	L_1	KIBMCE	0.9885	0.0561	0.0480	0.0471	0.0589	0.0510
		KMCE		0.0659	0.0530	0.0507	0.0700	0.0598
	L_2	KIBMCE	0.9885	0.0405	0.0362	0.0381	0.0482	0.0302
		KMCE		0.0481	0.0389	0.0393	0.0582	0.0339
	WISE	KIBMCE	0.9885	0.0404	0.0359	0.0402	0.0378	0.0408
		KMCE		0.0481	0.0384	0.0417	0.0450	0.0501

Table 4: *Ratio between the error measures of KIBMCE and KMCE.*

		Lognormal	Log-Pareto		Weibull	Tr. Logist.
			$p=0.7$	$p=0.3$		
$N=100$	L_1	0.9888	0.9637	0.9725	0.9982	0.9629
	L_2	0.9563	1.0162	1.0190	1.0100	0.9398
	WISE	0.9470	0.9538	0.9350	0.9835	0.9916
$N=1000$	L_1	0.8517	0.9059	0.9295	0.8413	0.8522
	L_2	0.8410	0.9295	0.9695	0.8284	0.8911
	WISE	0.8390	0.9340	0.9637	0.8392	0.8150

$l=0.99$ and $l=0.98$ with Buch-Larsen method and shows that double-transformation method improves the results presented in Buch-Larsen *et al.* (2005). In this work, we compare the results obtained when using an optimal trimming parameter l_{a^*} for $B(3,3)$.

We measure the performance of the estimators by the error measures based in L_1 norm, L_2 norm and *WISE*. This last weighs the distance between the estimated and the true distribution with the squared value of x. This results in an error measure that emphasizes the tail of the distribution:

$$\left(\int_0^\infty \left(\widehat{f}(x) - f(x) \right)^2 x^2 \, dx \right)^{1/2}.$$

The simulation results can be found in Table 3. For every simulated density and for sample sizes $N = 100$ and $N = 1000$, the results presented here correspond to the following error measures L_1, L_2 and $WISE$. The benchmark results are labeled KMCE and they correspond to those presented in Buch-Larsen *et al.* (2005). In Table 4 we show ratios between the error measures of KIBMCE and KMCE, if this ratio is smaller than 1 then KIBMCE improves on the results of KMCE.

In Table 4 we show that for $N = 100$ the ratios associated to L_1 and $WISE$ are always below one, and this indicates the KIBMCE method improves the fit of the density in the tail values of the density even for a small sample size. When $N = 1000$ then KIBMCE has always smaller values than KMCE, both for L_1, L_2 and for $WISE$. The best results can be obtained for the Lognormal, the Weibull and the Truncated Logistic, where in all cases the errors of KMCE are reduced by more that a 10% when using the new method.

For the mixtures of a Lognormal and a Pareto the results show that when $p = .7$ (70% Lognormal) the improvement is almost 10% for L_1 and is around 7% for L_2 and $WISE$. When $p = .3$ (30% Lognormal) L_1 is reduced by 7%, and both L_2 and $WISE$ are reduced in slightly more than 3%.

4. Data analysis

In this section, we apply our estimation method to a data set that contains automobile claim costs from a Spanish insurance company for accidents occurred in 1997. It is a typical insurance cost of individual claims data set, i.e. a large sample that looks heavy-tailed. The data are divided into two age groups: claims from policyholders who are less than 30 years old, and claims from policyholders who are 30 years old or older. The first group consists of 1,061 observations in the interval [1;126,000] with mean value 402.70. The second group contains 4,061 observations in the interval [1;17,000] with mean value 243.09. Estimation of the parameters in the modified Champernowne distribution function for the two samples of is, for young drivers $\widehat{\alpha}_1 = 1.116$, $\widehat{M}_1 = 66$, $\widehat{c}_1 = 0.000$ and for older drivers $\widehat{\alpha}_2 = 1.145$, $\widehat{M}_2 = 68$, $\widehat{c}_2 = 0.000$, respectively. We notice that $\alpha_1 < \alpha_2$, which indicates that the data set for young drivers has a heavier tail than the data set for older drivers.

To produce the graphics, the claims have been split into three categories: *Small claims* in the interval (0; 2,000), *moderately sized claims* in the interval [2,000; 14,000), and *extreme claims* in the interval [14,000; ∞). In Figure 3 we show the density function in the three categories for younger and older drivers.

Figure 3 shows that the KIBMCE method corrects the results of KMCE. In general, the density that is estimated using a KIBMCE method is smoother and larger in the mode, if compared with the KMCE estimate. In the tail and compared to the KMCE, the KIBMCE estimates a larger density in the tail, when the tails heavier, as for younger drivers, and it also estimates a smaller density when the tail is lighter, as for older drivers.

a) Younger drivers b) Older drivers

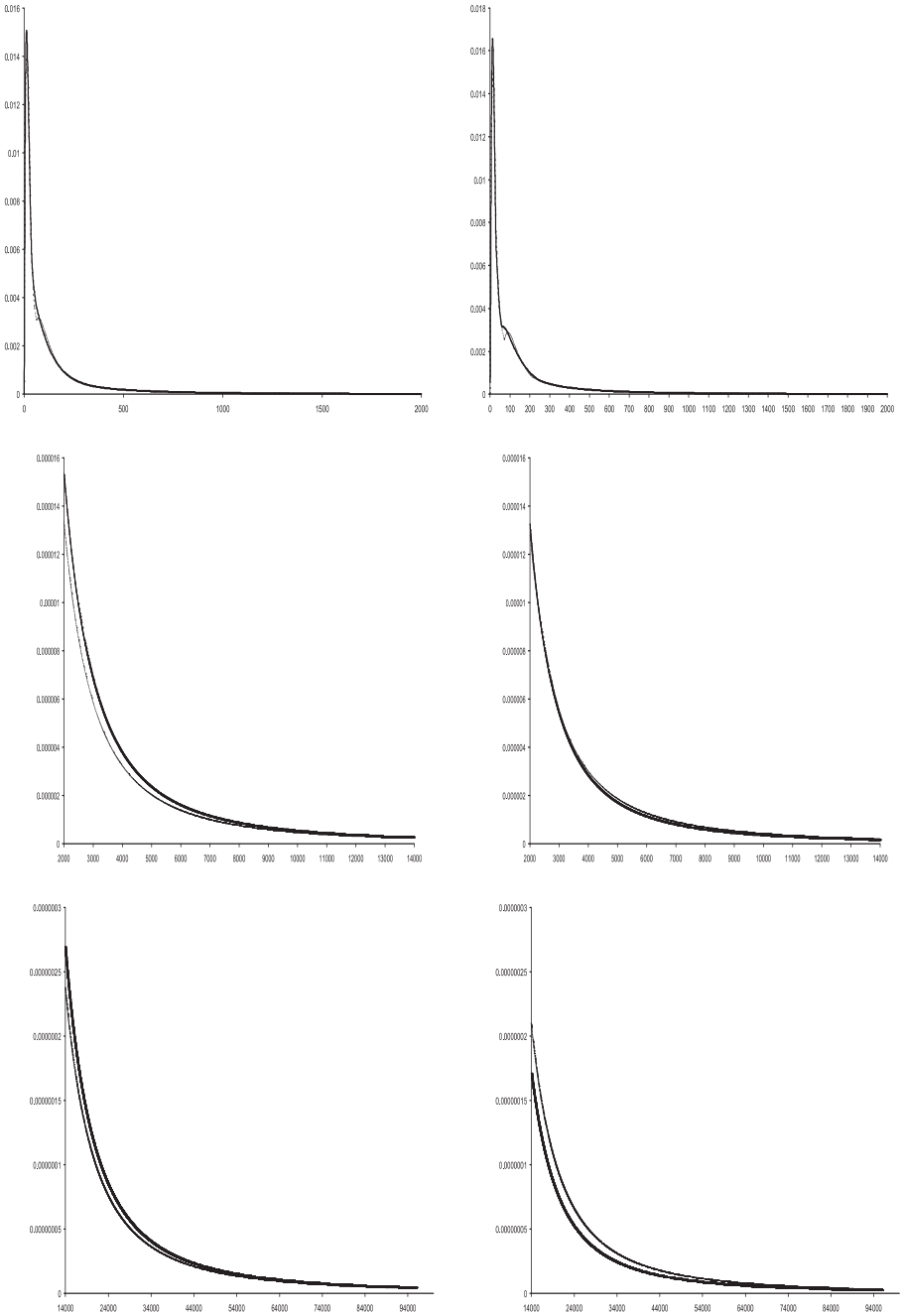


Figure 3: *Optimal KIBMCE estimates (thick) versus KMCE estimates (light) of insurance claims cost densities. Upper plots show small claims, middle plots show moderate claims and lower plots show large claims.*

5. Conclusions

In this work we have proposed a transformation kernel density estimator that can provide good results when the density to be estimated is very asymmetric and has extreme values. Moreover, the method presented here has a very straightforward method to calculate the smoothing parameter. This method provides a rule of thumb method to calculate the bandwidth in the context of transformation kernel density estimation that is comparable to Silverman's rule of thumb in the context of classical kernel density estimation. For large sample sizes, like the ones shown in the application, the simulation study shows that this method outperforms existing alternatives.

References

Bolancé, C., Guillén, M. and Nielsen, J. P. (2009). Transformation kernel estimation of insurance claim cost distribution, in Corazza, M. and Pizzi, C. (Eds). *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, Springer, 223-231.

Bolancé, C., Guillén, M. and Nielsen, J. P. (2008). Inverse Beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78, 1757-1764.

Bolancé, C., Guillén, M. and Nielsen, J. P. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, 32, 19-36.

Bolancé, C., Guillén, M., Pelican, E. and Vernic, R. (2008). Skewed bivariate models and nonparametric estimation for CTE risk measure. *Insurance: Mathematics and Economics*, 43, 386-393.

Buch-Larsen, T., Guillen, M., Nielsen, J. P. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39, 503-518.

Clements, A. E., Hurn, A. S. and Lindsay, K. A. (2003). Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, 98, 993-1000.

Degen, M. and Embrechts, P. (2008). EVT-based estimation of risk capital and convergence of high quantiles. *Advances in Applied Probability*, 40, 696-715.

Jones, M. C. (1990). Variable kernel density estimation and variable kernel density estimation. *Australian Journal of Statistics*, 32, 361-371.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, 85, 270-277.

Terrell, G. R. and Scott, D. W. (1985). Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, 80, 209-214.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* London: Chapman & Hall.

Wand, P., Marron, J. S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the Americal Statistical Association*, 86, 414, 343-361.

Selected article from

XII Conferencia Española de Biometría 2009

SORT 34 (2) July-December 2010, 239-248

Application of receiver operating characteristic (ROC) methodology in biological studies on marine resources: sex determination of *Paracentrotus lividus* (Lamarck, 1816)

Vicente Lustres-Pérez[1], María Xosé Rodríguez-Álvarez[2,3], María P. Pata[1]
Eugenio Fernández-Pulpeiro[1], Carmen Cadarso-Suárez[2,3]
Universidade de Santiago de Compostela (USC)

Abstract

The receiver operating characteristic (ROC) curve is usually used in biomedicine as an indicator of the accuracy of diagnostic tests. However, this measure of discrimination has been little used in other areas, such as animal biology or ecology. We present a novel application of an ROC analysis in which gonad colour was used to determine the sex of *Paracentrotus lividus* (Lamarck, 1816), a sea urchin of considerable commercial interest. A better classifier than gonad colour was obtained by transforming these colours through flexible logistic generalized additive models.

MSC: 6207, 62G08, 62G09, 62H30

Keywords: ROC, GAM, *Paracentrotus lividus*, bootstrap

1. Introduction

Paracentrotus lividus (Lamarck, 1816) is an echinoderm of high commercial value that is found along the coasts of Europe and North Africa. As this species is particularly abundant on the coast of Galicia (NW Spain), commercial harvesting began in the early 1980s. Although the reported annual average catch is in the order of 750 T (as per official

[1] Departamento de Zoología y Antropología Física, Universidade de Santiago de Compostela (USC), Spain.
[2] Unidad de Bioestadística, Departamento de Estadística e Investigación Operativa, USC, Spain.
[3] Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain.
Received: November 2009
Accepted: June 2010

data published by the Galician Regional Authority/*Xunta de Galicia*: www.pescadegalicia.com), real production is in fact higher, as a large volume of the catch goes undeclared. Currently, the *P. lividus* harvesting period in Galicia lasts from October to April, coinciding with the time of year when this species reaches sexual maturity. The spawning period on this stretch of coast usually begins at some point between the start and middle of spring (Catoira, 1995; Monteiro-Torreiro and Garcia-Martinez, 2003; Lustres-Pérez, 2006).

Although *P. lividus* is a dioecious species, with separate sexes, studies conducted by other authors report no method of determining the creature's sex externally, despite its outward display of a great variety of colours (Tortonese, 1965). Nevertheless, gonad colour has been linked to sex by some authors, though the majority of such studies have been of a descriptive nature (see e.g. Sellem and Guillou, 2007).

Commercial interest in this species lies in exploitation of the reproductive organs, the gonads. Female gonads are of higher quality than those of males, inasmuch as the former are reputed to have a better flavour and so be more palatable. This, in turn, means that clear criteria for sexual selection are vital for proper commercial and biological management.

From a statistical point of view, the discriminatory capacity of a given continuous or ordinal classifier, *Y* (gonad colour in our case), in terms of distinguishing between two alternative states, S1 and S2 (i.e., sex), is usually based on receiver operating characteristic (ROC) curve analysis (Metz, 1978; Swets and Pickett, 1982; Hanley and McNeil, 1982). The ROC curve is based on dichotomisation of the classifier *Y* by choosing a cut-off, such that values above this value will classify an individual as belonging to one of the states (say S1), and values below it as belonging to the alternative state (S2).

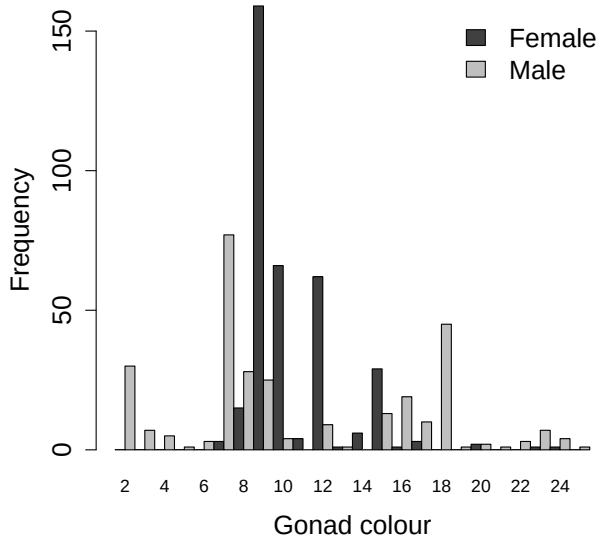


Figure 1: *Frequency of the different gonad colours selected according to sex.*

In many situations, however, the classification rule based on Y values that minimise the overall misclassification error is not necessarily the criterion used in ROC analysis. Figure 1 shows gonad colour frequency (codified numerically) for each sex. An irregular distribution of gonad colour can be observed, with a dominance of sexes in non-contiguous regions. Consequently, sex classification by means of a cut-off value is neither feasible nor logical. A modification of the classification rule is thus necessary. Indeed, if the discriminatory capacity of gonad colour is evaluated by means of the ROC curve, a lower discriminatory capacity will be obtained than that to be expected from a visual examination of Figure 1 (see Results, Section 3). Hence, use of such an analysis would lead to erroneous conclusions.

An intuitive solution to this problem, is to estimate the probability of belonging to one of the states as a function of the values of the marker Y (e.g., $P[S1|Y]$), and to base the classification on these probabilities, i.e., to transform the marker in such a way that the classification rules can be based on cut-off values.

This study proposes to model $P[S1|Y]$ by means of a generalised additive model (GAM, Hastie and Tibshirani, 1990) for binary data. GAMs are flexible non-parametric regression models that allow for much more accurate fitting of real data than do the parametric linear models usually used. Furthermore, in the case of the *P. lividus* data shown above, the use of a generalised linear model (GLM, McCullagh and Nelder, 1989) would not enable this probability to be correct modelled (see Section 3 for more details).

This paper is structured as follows: Section 2 outlines the statistical methodology; Section 3 reports the results of applying the proposed methodology to *P. lividus* data; and lastly, Section 4 concludes with a discussion.

2. Statistical methodology

Let Y be a continuous or ordinal classifier. Classification on the basis of Y of an individual as belonging to state S1 or S2 can be made by choosing a cut-off value, c, such that if $Y \geq c$ the observation is classified as S1, and if $Y < c$ it is classified as S2. Hence, each cut-off value chosen, c, will give rise to a true positive fraction (or sensitivity), $TPF(c) = P[Y \geq c|S1]$, and a false positive fraction (or 1-specificity), $FPF(c) = P[Y < c|S2]$. In such a situation, the ROC curve is defined as the set of all TPF-FPF pairs that can be obtained by a varying cut-off value c, $\{(FPF(c),\ TPF(c)),\ c \in (-\infty, \infty)\}$, or, equivalently, as the function of the form $ROC(t) = S_{S1}\left(S_{S2}^{-1}(t)\right)\ t \in (0,1)$, where S_{S2} and S_{S1} denote the survival functions of Y in the groups defined by states S1 and S2 respectively. Several indices can be used as summaries of the discriminatory capacity of the ROC curve. The area under the ROC curve (AUC) is most commonly used, taking values from 0.5 (no power of discrimination) to 1 (perfect power of discrimination). Generally, discrimination is deemed accurate where AUC exceeds 0.8.

As was illustrated in the Introduction with real *P. lividus* data, in many situations the classification rule based on the classifier Y that minimises the overall misclassification

error, is not the criterion used in ROC analysis. Moreover, the best Y-based classifier with a cut-off as the classification decision is that which is based on the conditional probability of one of the states (e.g., S1) given the values of Y (Neyman and Pearson, 1933; McIntosh and Pepe, 2002). Therefore, the best classifier, $\tilde{Y}$, can be expressed as:

$$\tilde{Y} \equiv f(Y) = P[S1|Y] \subset (0,1). \tag{1}$$

In practice, however, the function $f(\cdot)$ of (1) is not known, and its estimation would be required. In this study, we propose to model the function $f(\cdot)$ using a logistic GAM regression model as follows:

$$f(Y) = P[S1|Y] = g^{-1}(\alpha + h(Y)) = \frac{\exp(\alpha + h(Y))}{1 + \exp(\alpha + h(Y))}, \tag{2}$$

where $g(\cdot)$ is the logit link function (known) and $h(\cdot)$ is a smooth unknown function.

To date, several approaches to estimating the model (2) have been suggested in the statistical literature, e.g., methods based on penalised regression splines (Eilers and Marx, 1996; Wood, 2003) or the Bayesian versions of these (Lang and Brezger, 2004). Alternatively, the local scoring algorithm with kernel-type smoothers can be also used (McCullagh and Nelder, 1989; Wand and Jones, 1995).

In this paper, penalised regression combined with thin plate splines as smoothers (Wood, 2003) are proposed for the purpose of estimating the function $h(\cdot)$. A crucial step in estimating $h(\cdot)$ is choosing the smoothing parameter that controls the smoothness of the resultant estimate. In this paper, the optimal smoothing parameter is chosen automatically by use of the Un-Biased Risk Estimator criterion (UBRE) (Wood, 2004).

Once the model (2) is fitted, the estimated probabilities are used as the new classifier, and the ROC curve and the corresponding AUC are obtained. In addition, bootstrap regression techniques (Efron and Tibshirani, 1993) are used to construct a 95% bootstrap confidence interval (CI) for the AUC.

3. Results

3.1. Materials and methods

The study was undertaken at the following two sites along Galicia's Atlantic seaboard (NW Spain): Punto Area das Vacas (42°06′54″ N; 008°54′30″ W) situated on the Vigo estuary (*Ría de Vigo*); and Lago (42°19′25″ N; 008°49′37″ W) located on Aldán Bay (*Ensenada de Aldán*), on the southern edge of the Pontevedra estuary (*Ría de Pontevedra*). Both sites are located in fishing area of *P. lividus* and feature extensive rocky areas with a high abundance of this specie. However, the sampling areas are exploited only occasionally.

Table 1: *List of colours selected (C, cyan; M, magenta; Y, yellow; and K, black).*

Code	Colour	Pantone CVC	%CMYK			
			C	M	Y	K
1	Bright yellow	*Yellow C*	0	0	100	0
2	Yellow	107	0	0	79	0
3	Pale yellow	100	0	0	51	0
4	Dark yellow	110	0	11	94	6
5	Yellow+Black	1405	0	38	100	65
6	Bright yellow orange	137	0	34	91	0
7	Orange yellow	136	0	27	79	0
8	Pale orange	1495	0	30	69	0
9	Orange	1505	0	38	76	0
10	Bright orange	*Orange* 021C	0	51	87	0
11	Orange pink	1485	0	23	56	0
12	Orange red	172	0	65	83	0
13	Red	*Warm red*	0	79	91	0
14	Dark red	1795	0	94	100	0
15	Dark orange	1595	0	65	100	9
16	Orange light brown	167	0	60	100	18
17	Orange brown	1605	0	56	100	30
18	Light brown orange	160	0	60	100	34
19	Light brown red	1815	0	91	100	51
20	Light brown	724	0	51	100	43
21	Brown orange	1615	0	56	100	43
22	Brown red	181	0	72	79	47
23	Brown	168	0	56	100	60
24	Dark brown	1545	0	51	100	83
25	Black	*Black C*	0	0	0	100

The species of algae present in the intertidal zone at both sites included *Lithophyllum incrustans* Philippi 1837, *Corallina officinalis* Linnaeus 1758, *Corallina elongata* J. Ellis & Solander 1786, *Chondrus crispus* Stackhouse 1797, *Bifurcaria bifurcata* R. Ross 1958, *Ulva rigida* C. Agardh 1823, etc., which are all very common along intertidal areas on the Galicia coast. For its part, there is a high abundance of *Saccorhiza polyschides* (Lightfoot) Batters 1902, in the sublittoral zone studied.

Samples were collected monthly from January 2002 to February 2003 along the lower intertidal zone of both sites, and along the sublittoral zone of the latter. Samples were collected randomly, with each comprising 25 individuals of *P. lividus*. A total of 750 specimens were finally studied.

The sex was determined according to the colour of gametic fluid. Histological examination of the gonads showed that male gonads emit white gametes, while female emission was orange in colour although some cases were red. These observations are in agreement with the findings of other studies (e.g. Crapp and Willis, 1975). Samples were disregarded where it was not possible to collect gametic fluid.

A colour table (Pantone CVC, Pantone Inc) was used to determine the gonad colour of the samples. The table breaks the colours down into four component parts, namely, cyan (C), magenta (M), yellow (Y) and black (K) (collectively, CMYK). A total of 25 colours were observed in the samples collected, and codified using Table 1. The observations were made by three researchers under constant low light conditions across the study.

3.2. Statistical modelling

The discriminatory capacity of gonad colour for distinguishing male from female individuals was assessed by using the following two different classifiers: (a) raw gonad colour (without transformation); and, (b) gonad colour transformed through equation (2). In the latter case, the following logistic GAM regression model was fitted:

$$f(Colour) = P[Sex = 1|Colour] = g^{-1}(\alpha + h(Colour)), \tag{3}$$

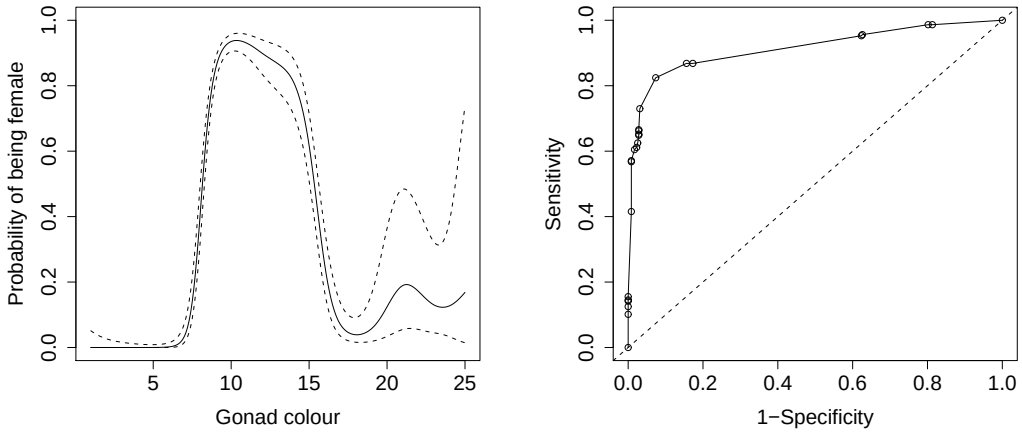
where *Colour* denotes gonad colour, *Sex* is a binary variable taking the value 1 for female and 0 for male, $g(\cdot)$ is the logit function (known), and $h(\cdot)$ is a smooth unknown function.

The logistic GAM regression model (3) was fitted by using the `gam` function of the `mgcv` package (Wood, 2006). The R software package, `ROCR` (Sing *et al.*, 2005), was used to estimate the ROC curve and AUC.

3.3. Results

The results shown below are based on the global data collected in the two sites of the study: Punto Area das Vacas and Lago. For the analysis of the discriminatory capacity of raw gonad colour (where the darkest values were assumed to be indicators of male gender), the estimated AUC was 0.586 with a 95% bootstrap confidence interval (CI) of (0.542, 0.638). Based on this result, gonad colour would not seem to be reliable for accurate classification of the sex of *P. lividus*.

With respect to the analysis performed with the transformed data, Figure 2a shows the estimated probability of being female according to gonad colour. Intermediate



(a) (b)

Figure 2: *(a) Estimated probability of being female according to gonad colour, together with its 95% confidence interval. Shown in (b) is the estimation of the ROC curve for gonad colour transformation.*

Table 2: *Range of colours for classifying an individual as female, along with the probabilities of reaching a correct decision.*

Range of female gonad colour	Probability of correct classification of a female individual (TPF)	Probability of correct classification of a male individual (TNF)
$[9, 14]$	0.84	0.87
$[9, 15]$	0.82	0.92

colours corresponded to a high probability of being female, while colours at the extremes of the palette corresponded to a low probability of being female and an ensuing higher probability of being male. The ROC curve associated with the above probability (employed as the classifier) is shown in Figure 2b. The corresponding AUC was 0.914 with a 95% bootstrap CI of (0.891, 0.936) ($n = 649$), with similar results being observed for all three populations after they had been separately analysed. Use of the logistic GAM regression model led to optimum predictive capacity. In contrast to the analysis of raw data, these results confirm that gonad colour can afford a high degree of accuracy in classifying the sex of *P. lividus*.

It is important to note that, on working with transformed data, the TPF-FPF pairs which give rise to the ROC curve are obtained on the basis of the probabilities estimated by the model (3). For this example, however, it is easy to obtain the gonad colour ranges that yield the said TPF-FPF pairs. Table 2 shows some possible colour ranges for classifying an individual as female, together with the corresponding probability of reaching a correct decision (in ROC terminology, the true positive, TPF, and the true negative, TNF, fractions).

4. Conclusions

In this paper, a new flexible alternative for evaluating the discriminatory capacity of a continuous or ordinal classifier is suggested. The proposed methodology is based on: (a) transformation of the classifier by means of a logistic GAM regression model; (b) use of the probabilities estimated by this model as a new classifier; and (c) evaluation of the discriminatory capacity of this new classifier by the ROC curve. This transformation makes it possible to obtain better cut-off values (or intervals) on which to base the classification.

The methodology presented in this paper was applied to the task of assessing the discriminatory capacity (accuracy) of gonad colour in terms of determining the sex of *P. lividus*. The results obtained with crude gonad colour indicated that this classifier had little accuracy. Yet when transformed gonad colour was used, this same discriminatory capacity proved to be high. Similarly, using transformed gonad colour, this paper furnishes two possible colour ranges on which to base *P. lividus* sex-classification in wild populations. Sexual determination of this species according to gonad colour serves both to enhance knowledge of its biology and to improve its commercial exploitation by enabling a better quality product to be obtained.

According to many authors, the diet of *P. lividus* greatly affects gonad colour, being particularly influenced by the accumulation of carotenoids (e.g. Shpigel *et al.*, 2005; Shpigel *et al.*, 2006). Many studies have investigated the influence of distinct diets (natural or artificial), in increasing gonad yield and improving gonad quality (from a commercial perspective). While the samples collected in this study originated from distinct habitats (intertidal/sublittoral), in which the diet of the urchins could be different, important changes were not observed neither in the distribution of gonad colours, nor in the ability of the gonad colour in discriminating the sex of *P. lividus*.

ROC methodology has an infinity of possibilities in the field of ecology and biology. In our opinion, the alternative presented in this paper could be of great utility in aspects relating to improvement in marine resource management, e.g., for determining the size or age at which examples of a species reach sexual maturity, or the periods during which a given resource reaches specific sexual stages. Likewise, it offers great possibilities in spatial distribution studies (presence/absence), among others. Application of this methodology will allow for solid results, based on appropriate statistical models, to be obtained.

Acknowledgements

The authors would like to express their gratitude for the support received in the form of the Spanish MEC Grant MTM2008-01603 and the Galician Regional Authority (Xunta de Galicia) projects INCITE08PXIB208113PR and 07MMA001200PR. We are also grateful to the referee for her/his valuable comments and suggestions, which served to make a substancial improvement to this paper.

References

Catoira, J. L. (1995). Spatial and temporal evolution of the gonad index of the sea urchin *Paracentrotus lividus* (Lamarck) in Galicia, Spain. In: Emson, R., Smith, A. and Campbell, A. (eds.). *Echinoderm Research*. Balkema, Rotterdam, 295-298.

Crapp, G. B. and Willis, M. E. (1975). Age determination in the sea urchin *Paracentrotus lividus* (Lamarck), with notes on the reproductive cycle. *Journal of Experimental Marine Biology and Ecology*, 20, 157-178.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRP Press, New York.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89-121.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Hastie, T. J and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall, London.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183-212.

Lustres-Pérez, V. (2006). El erizo de mar: *Paracentrotus lividus* (Lamarck, 1816) en las costas de Galicia. PhD thesis. Universidad de Santiago de Compostela.

McIntosh, M. W. and Pepe, M. S. (2002). Combining several screening tests: optimality of the risk score. *Biometrics*, 58, 657-664.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Second Edition. Chapman and Hall, London.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8, 283-298.

Monteiro-Torreiro, M. F. and Garcia-Martinez, P. (2003). Seasonal changes in the biochemical composition of body components of the sea urchin, *Paracentrotus lividus*, in Lorbé (Galicia-north-western Spain). *Journal of the Marine Biological Association of the United Kingdom*, 83, 575-581.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypothesis. *Philosophical Transactions of the Royal Society of London, Series A*, 231, 289-337.

Sellem, F. and Guillou, M. (2007). Reproductive biology of *Paracentrotus lividus* (Echinodermata: Echinoidea) in two contrasting habitats of northern Tunisia (south-east Mediterranean). *Journal of the Marine Biological Association of the United Kingdom*, 87, 763-767.

Shpigel, M., McBride, S. C., Marciano, S., Ron, S. and Ben-Amotz, A. (2005). Improving gonad colour and somatic index in the European sea urchin *Paracentrotus lividus*. *Aquaculture*, 245, 101-109.

Shpigel, M., Schlosser, S. C., Ben-Amotz, A., Lawrence, A. L. and Lawrence, J. M. (2006). Effects of dietary carotenoid on the gut and the gonad of the sea urchin *Paracentrotus lividus*. *Aquaculture*, 261, 1269-1280.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20), 3940-3941.

Swets, J. A. and Pickett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.

Tortonese, E. (1965). *Fauna d'Italia. Echinodermata*. Calderini, Bologna.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman & Hall, London.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B* 65, 95-114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99, 673-686.

Wood, S. N. (2006). *Generalized Additive Models, An Introduction with R.* Chapman & Hall/CRC, Boca Raton, Florida.

Xunta de Galicia. Plataforma tecnolóxica da pesca. Consellería do mar. http://www.pescadegalicia.com/default.htm (accessed: 23 November, 2010).

Book review

Biplots in Practice

Michael Greenacre

BBVA Foundation, Rubes Editorial

Biplots in Practice is, as the title states, evidently a book about biplots. The book has a very didactic format, with short chapters giving some theory and examples, followed by a summary of the main points of the chapter, a style that is strongly reminiscent of the author's earlier book, *Correspondence Analysis in Practice* (Greenacre, 1993). The book is aimed at applied scientists who have a need to convert large tables of numbers into graphical displays, though will be useful for students in multivariate analysis as well. For a good understanding of the text, some background in matrix algebra and regression is required. Each chapter of the book basically presents a particular type of biplot, related to a specific multivariate technique. The final three chapters concern case studies in biomedicine (gene expression data), socioeconomics (survey research) and ecology (fish morphology and diet). The book has four appendices: a computational appendix with the R code, a bibliography on biplot literature, a glossary of terms and an epilogue by the author. The book is available in electronic format on-line at the website of the BBVA foundation at no cost. On-line books offer the possibility of continued correction and modification which potentially may convert this book into a "living book". The graphics and typesetting of the book are excellent, it is very difficult to find any mistakes in text or formulas.

The book introduces the biplot in a very elegant manner: as a multivariate generalization of the scatterplot, linked to the factorization of a data matrix as the product of two matrices: the biplot points and the biplot vectors. The definition of the scalar product and it associated geometry then follow naturally. Calibration of the biplot vectors is used to illustrate biplot interpretation.

Chapters 2, 3 and 4 link biplots to trivariate regression, using the analogy between the regression factorization $\hat{\mathbf{Y}} = \mathbf{X}\,\mathbf{B}$ and the data matrix factorization $\mathbf{X} = \mathbf{A}\,\mathbf{B}'$ used in biplots. Biplot vectors are presented as gradient vectors of a plane in three dimensions that point towards the direction of steepest ascent. Calibration is again used to equip the gradient vectors with scales, to show that the gradient vector hardly differs from an ordinary scatterplot axis. The idea is extended to regressions with transformed response variables.

The book has a nice strong focus on the close link between biplots and regression. Biplot coordinates (both points and vectors) can always be interpreted as regression coefficients. Once this relationship is understood, it becomes a particularly easy exercise to fit supplementary points and supplementary variables in a biplot, indeed, by just doing regression and plotting the regression coefficients. The extensive terminology (GLM biplot, poisson regression biplot, logistic regression biplot, MDS biplot, etc.) introduced in these chapters seems superfluous. It suggests we have many "different" biplots, but in fact we use the same regression principle all the time, and the single term "regression biplot" would suffice. Another point is the geometric framework in which these chapters are cast. The reader has to imagine a plane in the third dimension, and plot the gradient vector into the horizontal plane below it. It may have appeal to many readers, but I think it is not necessary to go to a third dimension. The biplot vector can also be found by searching for an optimal direction for a variable *within* the two-dimensional scatterplot of the predictors. Least squares minimization of projection errors obtained when projecting scatterplot points onto vectors inside the scatterplot will lead directly to the regression formula for representing the variable (Graffelman & Aluja-Banet (2003)). In fact, the term "supplementary variable" is sorely missing in chapters 2 through 4: the truth is that we are trying to fit supplementary variables in two-dimensional scatterplots all the time.

Chapter 5 tackles what, from a didactical point of view, is probably the most challenging part of biplot theory: the singular value decomposition (SVD). Depending on the public, a lecturer in statistics may wish to explain biplots without the SVD, and precisely the previous chapters of the book have shown that this is very well possible. However, if the audience has a basic understanding of matrix algebra, then the SVD is certainly enlightening as the unifying matrix approximation tool underlying many multivariate methods, and it will pave the way for explaining row and column coordinates, goodness of fit, and differences in scaling. The exposition of the SVD in this chapter is neat and concise, rank and dimensionality are smoothly presented, and the author proceeds from the unweighted to the weighted case, with both weights for cases and variables. The last sections of the chapter treat the approximation of a symmetric distance matrix by the SVD, to show the link between PCA and classical (metric) scaling. I feel that this section will not be understood by readers who do not have a solid background in multidimensional scaling (MDS), as the double-centring of the distance matrix and the multiplication by $-\frac{1}{2}$ are left unexplained.

The next chapter on biplots in principal component analysis (PCA) is in my eyes the most controversial chapter of the book. First of all, the notation used here for PCA is far from standard. PCA is mainly used for analyzing a quantitative data matrix, and it is fairly standard to refer to the latter as a $n \times p$ matrix (cases times variables) instead of the I times J employed by the author. Then, the centred and scaled data matrix is called $\mathbf{S}$, whereas as $\mathbf{S}$ is the typical notation used to indicate a covariance matrix. Finally,

to indicate row and column coordinates four matrices are used, $\mathbf{F}$, $\mathbf{\Gamma}$ and $\mathbf{G}$, $\mathbf{\Psi}$. Since $\mathbf{F}$ and $\mathbf{\Gamma}$ refer to the same entities (rows), they are better indicated by the same letter, and using a different subscript to indicate the scaling. The same applies to the column markers. One cannot escape from the impression that good old PCA is dressed up and put on stage in a correspondence analysis outfit, using the author's notation from the latter context.

Curiously enough, the term "principal components" seems mainly restricted to the title of the chapter, the components are not mentioned, computed or interpreted. I would recommend computing and plotting the principal components, in order to make these new synthetic variables tangible. Moreover, a scatterplot of the principal components is half a biplot, only the arrows for the variables are missing to complete the latter. Matrix $\mathbf{F}$ in this chapter comes close to the principal components: it contains the components but divided by a factor $\sqrt{p}$. Why so? The fact that we obtained scaled principal components is a direct consequence of scaling the matrix that enters the SVD by $1/\sqrt{p}$. Consequently, the singular values are scaled by $\sqrt{p}$, and the eigenvalues by p. It may be a matter of taste ("cada maestrillo tiene su librillo" as they say in Spain), but I'd rather prefer the SVD of $(1/\sqrt{n})\mathbf{X}_c$, where $\mathbf{X}_c$ contains the centred data. This way the SVD takes "half" of the expression of the covariance matrix, and the squared singular values are the eigenvalues of the covariance matrix and also the variances of the principal components. Most of the things we compute then have a direct interpretation, interpretations that are lost in the rescaling used in the book. The fact that the eigenvalues in the book are the eigenvalues of the covariance matrix but divided by p, provokes that all eigenvalues are smaller, and that the differences between the successive eigenvalues become smaller as well. Consequently, the usual difference in dispersion between the horizontal and vertical axis in the PCA biplot becomes attenuated, more difficult to perceive. If you teach PCA by maximizing the variance of a linear combination of the variables, then it is nice to be able to show plots where the higher variance of the first component is clearly visible. The rescaling used in the book obscures this. A very positive aspect of this chapter is that it presents the full variance decomposition over axes and over points, showing the computation of goodness of fit for each point, and contributions to axes. These additional statistics have always accompanied standard CA output, but were rarely computed in PCA. Another point is that PCA biplots in this book are all based on a PCA of the covariance matrix. A different type of biplot is possible by doing a PCA of the correlation matrix. There are no simple linear relationships that relate the results of covariance based PCA and a correlation based PCA. The latter may actually be the more common form of PCA, because it is often used when the variables have different units. In biplots from a correlation based PCA scalar products between vectors approximate the correlations between the variables. A full treatment of PCA biplots then requires four biplots: two for the covariance based PCA and two for the correlation based PCA, with the singular values to the right or to the left in each case.

Chapter 7 is an interesting contribution, showing how data that have been transformed as log-ratios can be represented in a biplot and interpreted, and how natural laws can be inferred from such plots.

The next three chapters deal with biplots in CA, moving from simple to multiple CA in a natural way: first comes a two-way table, then concatenated tables, and finally the full Burt matrix. The first CA chapter starts with a controversial phrase "CA is the most versatile of the methods based on the SVD for visualizing data". Metric multidimensional scaling (in a weighted form), also known as principal coordinate analysis, underlies CA and many other multivariate methods and may therefore be regarded more versatile. Classical canonical *correlation* analysis (CCO), (not to be confused with canonical *correspondence* analysis (CCA)) also underlies CA, and may also be considered more versatile. Canonical correlation analysis allows the construction of biplots of the between set correlation matrix (Haber and Gabriel, 1976; Ter Braak, 1990; Graffelman, 2005). These biplots are not treated in this book, and that may be considered an omission, since these are tightly related to the CA biplots described in the book.

Simple CA is concisely presented by means of the SVD of the matrix of standardized residuals. For the unfamiliarized, the "standardized residuals" may fall a bit out of the sky, for why would we want to analyze standardized residuals? Some indications that CA studies deviations from an independence model would be welcome in this context. The asymmetric CA biplots are presented with examples. The final section on CA presents the "contribution biplot", a rescaled version of an asymmetric biplot that allows us to easily identify the main contributors to each axis. But is this contribution biplot now really the most interesting way to communicate the results? When interpreting a biplot, we may rather like to focus on those points that have high goodness of fit, so that we are safe about our interpretations. Thus, why don't we scale the standard coordinates in such a way that their vector length equals R^2 of the corresponding regression? This way the longest vectors correspond to the best represented column categories, and they are easily identified as such. It can all be done, and we call the corresponding biplot a "quality biplot", and another biplot scaling is born. It's not my purpose to create new biplot scalings, I raise this issue because in my opinion statisticians have proposed so many ways of scaling biplots that the situation has become chaotic. An inexperienced researcher wishing to make some biplots is confronted with a myriad of scaling possibilities, and will have a hard time just to figure out which scaling is needed, and wondering whether he/she has chosen the "right" scaling for his/her dataset, and be pretty much upset by the fact that the plots resulting from different scalings can look rather different. I feel that for the users of biplot methodology, some simple practical rules are needed, but it is beyond the scope of this review to expose them here in detail. Representing supplementary points in biplots, a classical issue in CA, is treated by using the weighted average relationship between rows and columns. This topic could

be very will linked with the regression approach from the first four chapters of the book, because the coordinates of a supplementary point in a biplot are regression coefficients. The regression approach is unifying, supplementary points in PCA can be obtained by applying the same principle.

The chapter on discriminant analysis biplots is less clear than the other chapters of the book. The topic is initially presented in close relationship with CA and log-ratio analysis, whereas the in the last section classical linear discriminant analysis (LDA) is presented in the form of a SVD. Biplots in LDA are not so well-known as PCA or CA biplots, which makes this chapter interesting. It seems more logical to treat the biplots obtained from classical LDA first. The author states that a CA of a set of concatenated tables is also a discriminant analysis, but this is far from clear, and not further explained.

Chapter 12 is an introduction to constrained biplots. The topic is presented from the perspective of the projection the data matrix of interest onto a subspace spanned by constraining variables.

The final three chapters are case studies demonstrating the use of biplot methodology in biomedicine (gene expression data), socioeconomics (survey research) and ecology (fish morphology and diet). Many of the classical texts in multivariate analysis still suffer from the fact that example data sets are analyzed that often do not even occupy half a printed page. The data sets used in this book, particularly those of the case studies, are of considerable size and come much closer to the large databases often used in modern research. Chapter 13 addresses the topic of reduction of the number of variables in a microarray experiment, with the purpose of identifying those variables (genes) that discriminate different types of cancer. The first section of the chapter tries to accomplish this by PCA, using sequential removal of genes based on the contribution to the PCA solution. This approach is open to a lot of methodological criticism. Why is contribution to the solution taken as a criterion? It is not specified how contribution is measured, is it with respect to a 2, 3, 4 or even higher dimensional solution? Moreover, in PCA there is no guarantee that the first few dimensions do contain the relevant information that separates the cancer types; part of this information may be present in the last principal component. The use of the procrustes statistic to monitor the change in the configuration also requires a choice of dimensionality that is left unspecified here. The final section repeats the analysis of the data, and is a very interesting application of the more natural approach, discriminant analysis, now taking contributions to group differentiation as a criterion for removal. Quadratic discriminant analysis is not considered. The second case study, chapter 14, contains applications of various form of CA to social survey data, with special attention for missing values and middle categories. The last case study investigates the relationships between two sets of variables registered for a sample of fish, morphological and diet variables. The author has chosen for constrained CA, and a constrained log-ratio analysis to analyze the data. The results are interesting, but there is ample margin for discussion of how these data should be analyzed. First of all, the

layout of the data, two sets of different, quantitative variables is the classical layout for canonical correlation analysis and for multivariate regression. So why not use these tools? Moreover, one of the main reasons for using the constrained approach (CCA or redundancy analysis (RDA)) in ecology is that there are typically more variables in one set than there are observations. This leads to singularity of the within set covariance matrix of one set of variables, and this inhibits the use of CCO, because it needs to invert these. But for the fish data, there are more fish than variables, and singularity is not a problem. CA is used for diet data with the argument that there are many zeros in the data set. However, the data come in percentage form. Some amalgamation of food categories may greatly reduce the number of zeros, and the CODA (compositional data analysis) approach of log-ratio transformations of the diet variables may become feasible. CCO or multivariate regression with two sets of log-ratio transformed variables then may be an alternative. A permutation test is used as a practical criterion for variable selection. However, if there are strong correlations between the predictors, the results may overstate the importance of the selected variables.

The computational appendix gives website references for downloading the data sets and R scripts. The scripts are documented in this appendix and show how to construct most of the biplots in the book. This will be of great practical value for the readers, enabling them to repeat or modify any analysis in the book, as well as for analyzing their own data.

Most of the literature on biplots is available in the form of research articles. The author has chosen not to include any references in the chapters, but to give some references with comments in an appendix. The same appendix also contains references to R software and R packages, and to some relevant websites. The bibliography does not pretend to be complete, though a few important references that are tightly related with some topics addressed in the book are missing: the seminal paper of Ter Braak (1986) on canonical (constrained) correspondence analysis, Gabriel and Odoroff (1990) and Graffelman and van Eeuwijk (2005) for the topic of biplot calibration. The LDA biplot in the book concerns an analysis of group means, and this is closely related to Gabriel's MANOVA biplot (Gabriel, 1995).

Finally, the epilogue gives additional reflections of the author about biplots and their future, and contains many useful recommendations on good biplot design beyond setting the aspect ratio to 1. *Biplots in Practice* is, in short, a very welcome text in the field that will certainly help to disseminate biplot theory and help many researchers to make nice pictures of their data.

Jan Graffelman
jan.graffelman@upc.edu
Department of Statistics and Operations Research
Universitat Politècnica de Catalunya

References

Gabriel, K. R. (1995). MANOVA biplots for two-way contingency tables. In Krzanowski, W. J., editor, *Recent Advances in Descriptive Multivariate Analysis*, 227-268.

Gabriel, K. R. and Odoroff, C. L. (1990). Biplots in biomedical research. *Statistics in Medicine*, 9, 469-485.

Graffelman, J. (2005). Enriched biplots for canonical correlation analysis. *Journal of Applied Statistics*, 32, 173-188.

Graffelman, J. and Aluja-Banet, T. (2003). Optimal representation of supplementary variables in biplots from principal component analysis and correspondence analysis. *Biometrical Journal*, 45, 491-509.

Graffelman, J. and van Eeuwijk, F. A. (2005). Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. *Biometrical Journal*, 47, 863-879.

Greenacre, M. J. (1993). *Correspondence Analysis in Practice*. Academic Press.

Haber, M. and Gabriel, K. R. (1976). Weighted least squares approximation of matrices and its application to canonical correlations and biplot display. Technical report, University of Rochester, Department of Statistics.

Ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167-1179.

Ter Braak, C. J. F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, 55, 519-531.

Information for authors and subscribers

Information for authors and subscribers

Submitting articles to SORT

Guidelines for submitting articles

SORT accepts for publication only original articles that have not been submitted to any other journal in the areas of Statistics, Operations Research, Official Statistics and Biometrics. Articles may be either theoretical or applied, and may include computational or educational elements. Publication will be exclusively in English, with an abstract in Catalan (abstract translation into Catalan can be done by the journal's staff).

All articles submitted to thematic sections will be forwarded for systematic peer review by independent specialists and/or members of the Editorial Board, except for those articles specifically invited by the journal or reprinted with permission. Reviewer comments will be sent to the primary author if changes are requested in form or content.

When an original article is received, the journal sends the author a dated receipt; this date will appear in the publication as "reception date". The date in which its final version is received will be the "acceptance date".

To submit an article, the author must send it in PDF format to the electronic address of the journal (sort@idescat.cat) especifying his e-mail address and/or a phone number. Once the article has been approved, the author must send a final electronic version, following the instructions of the editor responsible for that article. It is recommended that this final version be submitted using a LaTeX 2_ε .

In any case, upon request the journal secretary will provide authors with LaTeX 2_ε templates and appropriate references to the MSC2000 classification of the American Mathematical Society. For more information about the criteria for using this classification system, authors may consult directly the Mathematics Subject Classification 2000 (MSC2000) at the website of the Institut d'Estadística de Catalunya (http://www.idescat.es/sort/Normes.stm).

Publishing rights and authors' opinions

Authored articles represent the opinions of the authors; the journal does not necessarily agree with the opinions expressed in authored articles.

Formal guidelines for articles

The first page of the article must contain the following items: title, name of the author(s), professional affiliation and complete address, and an abstract (100-200 words) followed by MSC2000 classification of the American Mathematical Society and the keywords. Before submitting an article, the author(s) would be well advised to ensure that the text uses correct English.

Bibliographic references have to be listed alphabetically at the end of the article, according to the following examples:

Citations
Mahalanobis (1936), Rao (1982b)

Journal articles
Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business & Economic Statistics*, 9, 73-84.

Books
Platek, R., Rao, J. N. K., Särndal, C. E. and Singh, M. P. (eds.) (1987). *Small Area Statistics: an International Symposium*. New York: John Wiley and Sons.

Parts of books
Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J.P. Klein and P.K. Goel (eds.), Kluwer Academic Publishers, 237-247.

Web files or "pages"
Nielsen, S. F. (2001). *Proper and improper multiple imputation*
http://www.stat.ku.dk/~feodor/publications/ (10th May 2003).

Multiple publications by a single author are to be listed chronologically. Explanatory notes should be numbered sequentially and placed at the bottom of the corresponding page. Tables and figures should also be numbered sequentially and will be reproduced directly from the submitted originals if it is impossible to include them in the electronic text.

SORT *(Statistics and Operations Research Transactions)*
Institut d'Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona (Spain)
Tel.: +34-93 412 09 24 or +34-93 412 00 88
Fax: +34-93 412 31 45
E-mail: sort@idescat.cat

How to cite articles published in SORT

Commenges, D. (2003). Likelihood for interval-censored observations from multi-state models. *SORT*, 27 (1), 1-12.

Subscription form
SORT *(Statistics and Operations Research Transactions)*

Name _____

_____

Organisation _____

_____

Street Address _____

_____

Zip/Postal code _____ City _____

State/Country _____ Tel. _____

Fax _____ NIF/VAT Registration Number _____

E-mail _____

Date _____

Signature

I wish to subscribe to **SORT *(Statistics and Operations Research Transactions)*** for the year 2008 (volume 32)

Annual subscription rates:

– Spain: €22 (4 % VAT included)
– Other countries: €25 (4 % VAT included)

Price for individual issues (current and back issues):

– Spain: €15/issue (4 % VAT included)
– Other countries: €17/issue (4 % VAT included)

Method of payment:

☐ Bank transfer to account number 2013-0100-53-0200698577

☐ Automatic bank withdrawal from the following account number

☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

☐ Check made payable to the Institut d´Estadística de Catalunya

Please send this subscription form (or a photocopy) to:

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45

Bank copy

Authorisation for automatic bank withdrawal in payment for

SORT *(Statistics and Operations Research Transactions)*

The undersigned _____

authorises Bank/Financial institution _____

located at (Street Address) _____

Zip/postal code _____ City _____

Country _____

to draft the subscription to **SORT *(Statistics and Operations Research Transactions)*** from my account

number ☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

Date _____

Signature

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45