ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal

Volume 6, No.3
November 2008

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
  Av. António José de Almeida, 2
  1000-043 LISBOA
  PORTUGAL
  Tel.: +351 218 426 100
  Fax: +351 218 426 364
  Web site: http://www.ine.pt
  Customer Support Service
  (National network): 808 201 808
  (Other networks): +351 226 050 748
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue …… **€ 10**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)… **€ 24**
- Annual subscription (No. 2, No. 3) …… **€ 16**

INDEX

REVSTAT – Statistical Journal
Volume 6, Number 3, November 2008, 199–230

SPECIAL CHARACTERIZATIONS OF STANDARD DISCRETE MODELS

Authors: Carlos Alberto de Bragança Pereira
– Department of Statistics, University of São Paulo, Brazil
cpereira@ime.usp.br

Julio Michael Stern
– Department of Applied Mathematics, University of São Paulo, Brazil
jstern@ime.usp.br

Received: June 2007 Revised: April 2008 Accepted: April 2008

Abstract:

- This article presents important properties of standard discrete distributions and its conjugate densities. The Bernoulli and Poisson processes are described as generators of such discrete models. A characterization of distributions by mixtures is also introduced.

  This article adopts a novel singular notation and representation. Singular representations are unusual in statistical texts. Nevertheless, the singular notation makes it simpler to extend and generalize theoretical results and greatly facilitates numerical and computational implementation.

Key-Words:

- *Characterization of discrete distribution; conjugate densities; singular representations.*

AMS Subject Classification:

- 62-02, 60E05, 62F15.

1. INTRODUCTION AND NOTATION

This article presents important properties of the distributions used for categorical data analysis. Regardless of the population size being known or unknown, or the specific observational stopping rule, the Bernoulli Processes generates the sampling distributions considered. On the other hand, the Gamma distribution generates the prior and posterior distributions obtained: Gamma, Gamma-Poisson, Dirichlet, and Dirichlet-Multinomial. The Poisson Processes as generator of sampling distributions is also considered.

The development of the theory in this article is self contained, seeking a unified treatment of a large variety of problems, including finite and infinite populations, contingency tables of arbitrary dimension, deficiently categorized data, logistic regressions, etc. These models also present a way of introducing non parametric solutions.

This article adopts a singular notation and representation, first used in Pereira and Stern (2005). Singular representations are unusual in statistical texts. Nevertheless, the singular notation makes it simpler to extend and generalize theoretical results and greatly facilitates numerical and computational implementation.

The generation form of the discrete sampling distributions presented in Section 2 is, in fact, a characterization method of such distributions. If one recalls that all the distribution classes being mixed are complete classes and are Blackwell sufficient for the Bernoulli processes, the mixing distributions are unique. This characterization method is completely described in Basu and Pereira (1983).

Section 9 describes the Reny–Aczel characterization of the Poisson distribution. Although it could be thought as a de Finetti type characterization this characterization is based on alternative requirements. While de Finetti characterization is based on a permutable infinite 0-1 process, Reny–Aczek characterization is based on a homogeneous Markov process in a finite interval, generating finite discrete Markov Chains. Using Reny–Aczel characterization, together with Theorem 3.1, one can obtain a characterization of Multinomial distributions.

Section 7 describes the Dirichlet of Second Kind. In this section we also show how to use a multivariate normal approximation to the logarithm of a random vector distributed as Dirichlet of Second Kind, and a log-normal approximation to a Gamma distribution, see Aitchison and Shen (1980). In many examples of the authors' consulting practice these approximations proved to be a powerful modeling tool, leading to efficient computational procedures.

Let us first define some matrix notation. The operator $f:s:t$, to be read *from* f *to* t with *step* s, indicates the vector $\left[f, f+s, f+2s, ..., t\right]$ or the corresponding index domain. $f:t$ is a short hand for $f:1:t$. Usually we write a matrix, A, with subscript row index and superscript column index. Hence, A_i^j is the element in the i-th row and j-th column of matrix A. Index vectors can be used to build a matrix by extracting from a larger matrix a given sub-set of rows and columns. For example, $A_{1:m/2}^{n/2:n}$ is the northeast block, i.e. the block with the first rows and last columns, from A. Alternatively, we may write a matrix with row and column indices in parenthesis. Hence, we may write the northeast block as $A(1:m/2, n/2:n)$. The next example shows a more general case of this notation:

$$A = \begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{bmatrix}, \qquad r = \begin{bmatrix} 1 & 3 \end{bmatrix}, \qquad s = \begin{bmatrix} 3 & 1 & 2 \end{bmatrix},$$

$$A_r^s = A(r,s) = \begin{bmatrix} 13 & 11 & 12 \\ 33 & 31 & 32 \end{bmatrix}.$$

$V > 0$ is a positive definite matrix. The Diagonal operator, diag, if applied to a square matrix, extracts the main diagonal as a vector, and if applied to a vector, produces the corresponding diagonal matrix:

$$\mathrm{diag}(A) = \begin{bmatrix} A_1^1 \\ A_2^2 \\ \vdots \\ A_n^n \end{bmatrix}, \qquad \mathrm{diag}(a) = \begin{bmatrix} a_1 & 0 & ... & 0 \\ 0 & a_2 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & a_n \end{bmatrix}.$$

A list of matrices can be indexed with left subscript or superscript indices. In case of block matrices, these left indices indicate the row and column block position, like in the following example:

$$A = \begin{bmatrix} {}_1^1A & {}_1^2A & ... & {}_1^sA \\ {}_2^1A & {}_2^2A & ... & {}_2^sA \\ \vdots & \vdots & \ddots & \vdots \\ {}_r^1A & {}_r^2A & ... & {}_r^sA \end{bmatrix}.$$

Hence, ${}_r^sA_i^j$ is the element in the i-th row and j-th column of the block situated at the r-th block of rows and s-th block of columns of matrix A. Alternatively, we may write block indices in braces, that is, we may write ${}_r^sA_i^j$ as $A\{r,s\}(i,j)$.

The Vec operator stacks the columns of the argument matrix in a single vector. The Kronecker product, also known as direct or tensor product, is defined

as follows:

$$\mathrm{Vec}(U^{1:n}) = \begin{bmatrix} u^1 \\ u^2 \\ \vdots \\ u^n \end{bmatrix}, \qquad A \otimes B = \begin{bmatrix} A_1^1 B & A_1^2 B & ... & A_1^n B \\ A_2^1 B & A_2^2 B & ... & A_2^n B \\ \vdots & \vdots & \ddots & \vdots \\ A_m^1 B & A_m^2 B & ... & A_m^n B \end{bmatrix}.$$

We now introduce some concepts and notations related to the permutation and partition of indices. Let $1:m$ be an index domain or, in this article context, a classification index. Let $p = \sigma(1:m)$ be a permutation of these indices. The corresponding (Row) Permutation Matrix is

$$P = I_p = \begin{bmatrix} I_{p(1)} \\ \vdots \\ I_{p(m)} \end{bmatrix}, \qquad \text{hence}, \qquad P \begin{bmatrix} 1 \\ \vdots \\ m \end{bmatrix} = \begin{bmatrix} p(1) \\ \vdots \\ p(m) \end{bmatrix}.$$

A permutation vector, p, and a termination vector, t, define a partition of the m original classes in s super-classes:

$$\begin{bmatrix} p(1) \\ \vdots \\ p\big(t(1)\big) \end{bmatrix}, \qquad \begin{bmatrix} p\big(t(1)+1\big) \\ \vdots \\ p\big(t(2)\big) \end{bmatrix}, \qquad ..., \qquad \begin{bmatrix} p\big(t(s-1)+1\big) \\ \vdots \\ p\big(t(s)\big) \end{bmatrix},$$

$$\text{where} \qquad t(0) = 0 < t(1) < ... < t(s-1) < t(s) = m\ .$$

We define the corresponding permutation and partition matrices, P and T, as

$$P = I_{p(1:m)} = \begin{bmatrix} {}_1P \\ {}_2P \\ \vdots \\ {}_sP \end{bmatrix}, \qquad {}_rP = I_{p\left(t(r-1)+1\,:\,t(r)\right)}\ ,$$

$$T_r = \mathbf{1}'({}_rP) \qquad \text{and} \qquad T = \begin{bmatrix} T_1 \\ \vdots \\ T_s \end{bmatrix}.$$

These matrices facilitate writing functions of a given partition, like

- The class indices in the super-class r

$${}_rP\,(1:m) \;=\; {}_rP \begin{bmatrix} 1 \\ \vdots \\ m \end{bmatrix} = \begin{bmatrix} p\big(t(r-1)+1\big) \\ \vdots \\ p\big(t(r)\big) \end{bmatrix};$$

- The number of classes in the super class r

$$T_r\,\mathbf{1} \;=\; t(r) - t(r-1)\ ;$$

- A sub-matrix with the row indices in super-class r

$$ {}_rP\,A \;=\; \begin{bmatrix} A_{p\left(t(r-1)+1\right)} \\ \vdots \\ A_{p\left(t(r)\right)} \end{bmatrix} ; $$

- The summation of the rows of a submatrix with row indices in super-class r

$$ T_r\,A \;=\; \mathbf{1}'({}_rP\,A)\,; $$

- The rows of a matrix, added over each super-class

$$ T\,A \;=\; \begin{bmatrix} T_1\,A \\ \vdots \\ T_s\,A \end{bmatrix} . $$

Note that a matrix T represents a partition of m-classes into s-super-classes if T has dimension $s \times m$, $T_h^j \in \{0,1\}$ and T has orthogonal rows. The element T_h^j indicates if the class $j \in 1:m$ is in super-class $h \in 1:s$.

We introduce the following notation for observation matrices, and respective summation vectors:

$$ U = \left[u^1, u^2, ...\right] , \qquad U^{1:n} = \left[u^1, u^2, ..., u^n\right] , \qquad x^n = U^{1:n}\,\mathbf{1} = \sum\nolimits_{j=1}^{n} u^j \ . $$

The tilde accent indicates some form of normalization like, for example, $\widetilde{x} = (1/\mathbf{1}'x)\,x$.

Lemma 1.1. *If $u^1, ..., u^n$ are i.i.d. random vectors,*

$$ x = U^{1:n}\,\mathbf{1} \implies \mathrm{E}(x) = n\,\mathrm{E}(u^1) \quad \textit{and} \quad \mathrm{Cov}(x) = n\,\mathrm{Cov}(u^1) \ . $$

Proof: The first result is trivial. For the second result, we only have to remember the transformation properties for the expectation and covariance operators by a linear operation on their argument,

$$ \mathrm{E}(AY+b) = A\,\mathrm{E}(Y) + b \ , \qquad \mathrm{Cov}(AY+b) = A\,\mathrm{Cov}(Y)\,A' \ , $$

and write

$$ \begin{aligned} \mathrm{Cov}(x) \;&=\; \mathrm{Cov}\big(U^{1:n}\,\mathbf{1}\big) \\ &=\; \mathrm{Cov}\Big(\big(\mathbf{1}'\otimes I\big)\,\mathrm{Vec}\big(U^{1:n}\big)\Big) \;=\; \big(\mathbf{1}'\otimes I\big)\,\big(I\otimes\mathrm{Cov}(u^1)\big)\,\big(\mathbf{1}\otimes I\big) \\ &=\; \big(\mathbf{1}'\otimes\mathrm{Cov}(u^1)\big)\,\big(\mathbf{1}\otimes I\big) \;=\; n\,\mathrm{Cov}(u^1) \ . \end{aligned} $$

□

2. THE BERNOULLI PROCESS

Let us consider a sequence of random vectors $u^1, u^2, ...$ where, $\forall u^i$ can assume only two values

$$I^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{or} \quad I^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \text{where} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

representing success or failure. That is, u^i can assume the value of any column of the identity matrix, I. We say that u^i is of class k, $c(u^i) = k$, iff $u^i = I^k$, $k \in [1, 2]$.

Also assume that (in your opinion), this sequence is exchangeable, that is, if $p = \big[p(1), p(2), ..., p(n)\big]$ is a permutation of $[1, 2, ..., n]$, then, $\forall\, n, p$,

$$\Pr\big(u^1, ..., u^n\big) \;=\; \Pr\big(u^{p(1)}, ..., u^{p(n)}\big)\;.$$

Just from this exchangeability constraint, that can be interpreted as saying that the index labels are non informative, de Finetti Theorem establishes the existence of an unknown vector

$$\theta \in \Theta \;=\; \left\{ \mathbf{0} \le \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \le \mathbf{1} \;\middle|\; \mathbf{1}'\theta = 1 \right\}$$

such that, conditionally on θ, $u^1, u^2, ...$ are mutually independent, and the conditional probability of $\Pr(u^i = I^k \,|\, \theta)$ is θ_k, i.e.

$$\big(u^1 \amalg u^2 \amalg ...\big) \,\big|\, \theta \quad \text{or} \quad \coprod_{i=1}^{\infty} u_i \,|\, \theta\;, \qquad \text{and} \qquad \Pr\big(u^i = I^k \,|\, \theta\big) = \theta_k\;.$$

Vector θ is characterized as the limit of proportions

$$\theta = \lim_{n\to\infty} \frac{1}{n}\, x^n\;, \qquad x^n = U^{1:n}\,\mathbf{1} = \sum\nolimits_{j=1}^{n} u^j\;.$$

Conditionally on θ, the sequence $u^1, u^2, ...$ receives the name of Bernoulli process. As we shall see, many well known discrete distributions can be obtained from transformations of this process.

The expectation and covariance (conditionally on θ) of any vector in the sequence are:

- $\mathrm{E}(u^i) = \theta\,;$
- $\mathrm{Cov}(u^i) = \mathrm{E}\big(u^i \otimes (u^i)'\big) - \mathrm{E}(u^i) \otimes \mathrm{E}\big((u^i)'\big) = \mathrm{diag}(\theta) - \theta \otimes \theta'\,.$

When the summation domain $1:n$ is understood, we may use the relaxed notation x instead of x^n. We also define the Delta operator, or "pointwise power

product" between two vectors of same dimension: Given θ, and x, $n \times 1$,

$$\theta \triangle x \equiv \prod_{i=1}^{n} (\theta_i)^{x_i} .$$

A stopping rule, δ, establishes, for every $n = 1, 2, ...$, a decision of observing (or not) u^{n+1}, after the observations $u^1, ..., u^n$.

For a good understanding of this text, it is necessary to have a clear interpretation of conditional expressions like $x^n \,|\, n$ or $x_2^n \,|\, x_1^n$. In both cases we are referring to a unknown vector, x^n, but with a different partial information. In the first case, we know n, and therefore we know the sum of components, $x_1^n + x_2^n = n$; however, we know neither component x_1^n nor x_2^n. In the second case we only know the first component, of x^n, x_1^n, and do not know the second component, x_2^n, obviously we also do not know the sum, $n = x_1^n + x_2^n$. Just pay attention: We list what we know to the right of the bar and, (unless we have some additional information) everything that can not be deduced from this list is unknown.

The first distribution we are going to discuss is the Binomial. Let $\delta(n)$ be the stopping rule where n is the pre-established number of observations. The (conditional) probability of the observation sequence $U^{1:n}$ is

$$\Pr\big(U^{1:n} \,|\, \theta\big) = \theta \triangle x^n .$$

The summation vector, x^n, has Binomial distribution with parameters n and θ, and we write $x^n \,|\, [n, \theta] \sim \text{Bi}(n, \theta)$. When n (or $\delta(n)$) is implicit in the context we may write $x \,|\, \theta$ instead of $x^n \,|\, [n, \theta]$. The Binomial distribution has the following expression:

$$\Pr\big(x^n \,|\, n, \theta\big) = \binom{n}{x^n} \big(\theta \triangle x^n\big)$$

where

$$\binom{n}{x} \equiv \frac{\Gamma(n+1)}{\Gamma(x_1+1)\,\Gamma(x_2+1)} = \frac{n!}{x_1!\; x_2!} \qquad \text{and} \qquad n = \mathbf{1}'x .$$

It is not hard to check that expectation vector and the covariance matrix of $x^n \,|\, [n, \theta]$ have the following expressions:

$$\text{E}(x^n) = n\theta \qquad \text{and} \qquad \text{Cov}(x^n) = n\,(\theta \triangle \mathbf{1}) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} .$$

The second distribution we discuss is the Negative Binomial. Let $\delta(x_1^n)$ be the rule establishing to stop at observation u^n when obtaining a pre-established

number of x_1^n successes. The random variable x_2^n, the number of failures he have when we obtain the required x_1^n successes, is called a Negative Binomial with parameters x_1^n and θ. It is not hard to prove that the Negative Binomial distribution $x_2^n \,|\, [x_1^n, \theta] \sim \mathrm{NB}(x_1^n, \theta)$, has expression, $\forall\, x_2^n \in \mathbb{N}$,

$$\Pr\big(x^n \,|\, x_1^n, \theta\big) \;=\; \frac{x_1^n}{n} \binom{n}{x^n} \big(\theta \triangle x^n\big) \;=\; \theta_1 \Pr\Big(\big(x^n - I^1\big) \,|\, (n-1),\, \theta\Big) \;.$$

Note that, from the definition of this distribution, x_1^n is a positive integer number. Nevertheless, we can extend the definition above for any real positive value a, and still obtain a probability function. For this, we use

$$\sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)\, j!} \,(1-\pi)^j \;=\; \pi^{-a} \;, \qquad \forall\, a \in [0, \infty[\quad \text{and} \quad \pi \in \,]0, 1[\;.$$

It is not hard to check the last equation, as well as the following expressions for the expectation and variance of x_2^n:

$$\mathrm{E}\big(x_2^n \,|\, x_1^n, \theta\big) = \frac{x_1^n\, \theta_2}{\theta_1} \qquad \text{and} \qquad \mathrm{Var}\big(x_2^n \,|\, x_1^n, \theta\big) = \frac{x_1^n\, \theta_2}{(\theta_1)^2} \;.$$

In the special case of $\delta(x_1^n = 1)$, the Negative Binomial distribution is also known as the Geometric distribution with parameter θ. If a random variables are independent and identically distributed (i.i.d.) as a geometric distribution with parameter θ, then the sum of these variables has Negative Binomial distribution with parameters a and θ.

The third distribution studied in this article is the Hypergeometric. Going back to the original sequence, $u^1, u^2, ...$, assume that a first observer knows the first N observations, while a second observer knows only a subsequence of $n < N$ of these observations. Since the original sequence, $u^1, u^2, ...$, is exchangeable, we can assume, without loss of generality, that the subsequence known to the second observer is the subsequence of the first n observations, $u^1, ..., u^n$. Using de Finetti theorem, we have that x^n and $x^N - x^n = U^{n+1\,:\,N}\mathbf{1}$ are conditionally independent, given θ. That is, $x^n \amalg (x^N - x^n) \,|\, \theta$. Moreover, we can write

$$x^n \,|\, [n, \theta] \sim \mathrm{Bi}(n, \theta) \;, \qquad x^N \,|\, [N, \theta] \sim \mathrm{Bi}(N, \theta) \qquad \text{and}$$

$$(x^N - x^n) \,\big|\, \big[(N-n), \theta\big] \;\sim\; \mathrm{Bi}(N-n, \theta) \;.$$

Our goal is to find the distribution function of $x^n \,|\, x^N$. Note that x^N is sufficient for $U^{1\,:\,N}$ given θ, and x^n is sufficient for $U^{1\,:\,n}$. Moreover $x^n \,|\, [n, x^N]$ has the same distribution of $x^n \,|\, [n, x^N, \theta]$. Using the basic rules of probability

calculus and the properties above, we have that

$$\begin{aligned}\Pr\big(x^n \,|\, n, x^N, \theta\big) &= \frac{\Pr\big(x^n, x^N | n, N, \theta\big)}{\Pr\big(x^N | n, N, \theta\big)} \\ &= \frac{\Pr\big(x^n, (x^N - x^n) \,|\, n, N, \theta\big)}{\Pr\big(x^N | n, N, \theta\big)} \\ &= \frac{\Pr\big(x^n \,|\, n, N, \theta\big)\ \Pr\big(x^N - x^n \,|\, n, N, \theta\big)}{\Pr\big(x^N | n, N, \theta\big)} \ .\end{aligned}$$

Hence, $x^n \,|\, [n, x^N]$ has distribution function

$$\Pr\big(x^n | n, x^N\big) = \frac{\binom{n}{x^n}\ \binom{N-n}{x^N - x^n}}{\binom{N}{x^N}}$$

$$\text{where} \qquad \mathbf{0} \le x^n \le x^N \le N\,\mathbf{1} \ , \qquad \mathbf{1}'x^n = n \ , \qquad \mathbf{1}'x^N = N \ .$$

This is the vector representation of the Hypergeometric probability distribution:

$$x^n \,|\, [n, x^N] \ \sim\ \mathrm{Hy}(n, N, x^N) \ .$$

It is not hard to check the following expressions for the expectation and (conditional) covariance of $x^n \,|\, [n, N, x^N]$, and covariance of u^i and u^j, $i, j \le n$:

$$\mathrm{E}(x^n) = \frac{n}{N}\, x^N \quad \text{and} \quad \mathrm{Cov}(x^n) \ = \ \frac{n(N-n)}{(N-1)}\ \big(x^N \triangle \mathbf{1}\big)\ \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ ,$$

$$\mathrm{Cov}(u^i, u^j \,|\, x^N) \ = \ \frac{1}{(N-1)\,N^2}\ \big(x^N \triangle \mathbf{1}\big)\ \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \ .$$

We finish this section presenting the derivation of the Beta-Binomial distribution. Let us assume that the first observer observed x_2^n failures, until observing a pre-established number of x_1^n successes. A second observer makes more observations, observing x_2^N failures until completing the pre-established number of x_1^N successes, $x_1^n < x_1^N$.

Since x_1^n and x_1^N are pre-established, we can write

$$x_2^N \,|\, \theta \ \sim\ \mathrm{NB}(x_1^N, \theta) \ , \qquad x_2^n \,|\, \theta \ \sim\ \mathrm{NB}(x_1^n, \theta) \ ,$$

$$(x_2^N - x_2^n) \,|\, \theta \ \sim\ \mathrm{NB}(x_1^N - x_1^n, \theta) \quad \text{and} \quad x_2^n \amalg (x_2^N - x_2^n) \,|\, \theta \ .$$

As before, our goal is to describe the distribution of $x_2^n \,|\, [x_1^n, x^N]$. If one notices that $[x_1^n, x^N]$ is sufficient for $[x^n, (x^N - x^n)]$, with respect to θ, the problem

becomes similar to the Hypergeometric case, and one can obtain

$$\Pr\big(x_2^n \,|\, x_1^n, x^N\big) \;=\; \frac{x_2^N!\,\Gamma(x_1^N)}{\Gamma(x_2^N + x_1^N)}\;\frac{\Gamma(x_2^n + x_1^n)}{x_2^n!\,\Gamma(x_1^n)}\;\frac{\Gamma\big(x_2^N - x_2^n + x_1^N - x_1^n\big)}{(x_2^N - x_2^n)!\,\Gamma(x_1^N - x_1^n)}\;,$$
$$x_2^n \in \big\{0, 1, ..., x_2^N\big\}\;.$$

This is the distribution function of a random variable called Beta Binomial with parameters x_1^n and x^N:

$$x_2^n \,|\, (x_1^n, x^N) \;\sim\; \mathrm{BB}(x_1^n, x^N)\;.$$

The properties of this distribution will be studied in the general case of the Dirichlet-Multinomial, in the following sections.

Generalized categories for $k > 2$ can be represented by the orthonormal base $I^1, I^2, ...I^k$, i.e., the columns of the k-dimensional identity matrix. The Multinomial and Hypergeometric multivariate distributions, presented in the next sections, are distributions derived of this basic generalization.

3. MULTINOMIAL DISTRIBUTION

Let u^i, $i = 1, 2, ...$, be random vectors with possible results in the set of columns of the m-dimensional identity matrix, I^k, $k \in 1:m$. We say that u^i is of class k, $c(u^i) = k$, iff $u^i = I^k$.

Let $\theta \in [0,1]^m$ be the vector of probabilities for an observation of class k in a m-variate Bernoulli process, i.e.,

$$\Pr\big(u^i = I^k \,|\, \theta\big) = \theta_k\;, \qquad \mathbf{0} \le \theta \le \mathbf{1}\,, \quad \mathbf{1}'\theta = 1\;.$$

Like in the last section, let U

$$U = [u^1, u^2, ...] \qquad \text{and} \qquad x^n = U^{1:n}\,\mathbf{1}\;.$$

Definition 3.1. If the knowledge of θ makes the vectors u^i independent, then the (conditional) distribution of x^n given θ is the Multinomial distribution of order m with parameters n and θ, given by

$$\Pr\big(x^n \,|\, n, \theta\big) \;=\; \binom{n}{x^n}\;(\theta \triangle x^n)$$

where

$$\binom{n}{x} \equiv \frac{\Gamma(n+1)}{\Gamma(x_1+1)\cdots\Gamma(x_m+1)} \;=\; \frac{n!}{x_1!\cdots x_m!} \qquad \text{and} \qquad n = \mathbf{1}'x\;.$$

We represent the m-Multinomial distribution writing

$$x^n \,|\, [n, \theta] \;\sim\; \mathrm{Mn}_m(n, \theta) \;.$$

When $m = 2$, we have the binomial case.

Let us now examine some properties of the Multinomial distribution.

Lemma 3.1. *If $x|\theta \sim \mathrm{Mn}_m(n, \theta)$ then the (conditional) expectation and covariance of x are*

$$\mathrm{E}(x) = n\,\theta \qquad \textit{and} \qquad \mathrm{Cov}(x) = n\Big(\mathrm{diag}(\theta) - \theta \otimes \theta'\Big) \;.$$

Proof: Analogous to the binomial case. □

The next result presents a characterization of the Multinomial in terms of the Poisson distribution.

Lemma 3.2. *Reproductive property of the Poisson distribution.*

$$x_i \sim \mathrm{Ps}(\lambda_i) \;\;\Longrightarrow\;\; \mathbf{1}'x \,|\, \lambda \;\sim\; \mathrm{Ps}(\mathbf{1}'\lambda) \;.$$

That is, the sum of (independent) Poisson variates is also Poisson.

Theorem 3.1. *Characterization of the Multinomial by the Poisson.*
Let $x = [x_1, ..., x_m]'$ be a vector with independent Poisson distributed components with parameters in the known vector $\lambda = [\lambda_1, ..., \lambda_m]' > 0$. Let n be a positive integer. Then, given λ,

$$x \,|\, [n = \mathbf{1}'x, \lambda] \;\sim\; \mathrm{Mn}_m(n, \theta) \qquad \textit{where} \;\; \theta = \frac{1}{\mathbf{1}'\lambda}\,\lambda \;.$$

Proof: The joint distribution of x, given λ is

$$\Pr(x \,|\, \lambda) \;=\; \prod_{k=1}^{m} \frac{e^{-\lambda_k}\lambda_i^{x_k}}{x_k!} \;.$$

Using the Poisson reproductive property,

$$\Pr\big(x \,|\, \mathbf{1}'x = n, \lambda\big) \;=\; \frac{\Pr\big(\mathbf{1}'x = n \wedge x \,|\, \lambda\big)}{\Pr\big(\mathbf{1}'x = n \,|\, \lambda\big)} \;=\; \delta(n = \mathbf{1}'x)\,\frac{\Pr(x \,|\, \lambda)}{\Pr\big(\mathbf{1}'x = n \,|\, \lambda\big)} \;. \qquad \square$$

The following results state important properties of the Multinomial distribution. The proof of these properties is simple, using the characterization of the Multinomial by the Poisson, and the Poisson reproductive property.

Theorem 3.2. *Multinomial Class Partition.*
Let $1:m$ be the index domain for the classes of a order m Multinomial distribution. Let T be a partition matrix breaking the m-classes into s-super-classes. Let $x \sim \mathrm{Mn}_m(n,\theta)$, then $y = Tx \sim \mathrm{Mn}_s(n, T\theta)$.

Theorem 3.3. *Multinomial Conditioning on the Partial Sum.*
If $x \sim \mathrm{Mn}_m(n,\theta)$, then the distribution of part of the vector x conditioned on its sum has Multinomial distribution, having as parameter the corresponding part of the original (normalized) parameters. In more detail, conditioning on the t first components, we have:

$$x_{1:t} \,|\, (\mathbf{1}' x_{1:t} = j) \;\sim\; \mathrm{Mn}_t\!\left(j\,,\; \frac{1}{\mathbf{1}'\theta_{1:t}}\,\theta_{1:t} \right) \qquad \text{where} \quad 0 \le j \le n\ .$$

Theorem 3.4. *Multinomial-Binomial Decomposition.*
Using the last two theorems (3.2 and 3.3), if $x \sim \mathrm{Mn}_m(n,\theta)$,

$$\begin{aligned}\Pr(x|n,\theta) \;=\; \sum_{j=0}^{n} & \Pr\!\left(x_{1:t} \,|\, j\,,\; \frac{1}{\mathbf{1}'\theta_{1:t}}\,\theta_{1:t} \right) \\ & \cdot\ \Pr\!\left(x_{t+1:m} \,|\, (n-j)\,,\; \frac{1}{\mathbf{1}'\theta_{t+1:m}}\,\theta_{t+1:m} \right) \\ & \cdot\ \Pr\!\left(\begin{bmatrix} j \\ (n-j) \end{bmatrix} \Big|\, n\,,\; \begin{bmatrix} \mathbf{1}'\theta_{1:t} \\ \mathbf{1}'\theta_{t+1:m} \end{bmatrix} \right)\ .\end{aligned}$$

Analogously, we could write the Multinomial-Trinomial decomposition for a three-partition of the class indices in three super-classes. More generally, we could also write the m-nomial-s-nomial decomposition for the partition of the m class indices into s super-classes.

4. MULTIVARIATE HYPERGEOMETRIC DISTRIBUTION

In the second section we have shown how an Hypergeometric variate can be generated from a Bernoulli process. The natural generalization of this result is obtained considering a Multinomial process. As in the last section, we say that u^i is of class k, $c(u^i) = k$, iff $u^i = I^k$.

We take a sample of size n from a finite population of size $N\,(> n)$, that is partitioned into m classes. The population frequencies (number of elements in each category) are represented by $[\psi_1, ..., \psi_m]$, hence $N = \mathbf{1}'\psi$. Based on the sample, we want to make an inference on ψ. x_k is the sample frequency of class k.

One way of describing this problem is to consider an urn with N balls of m different colors, indexed by $1, ..., m$. ψ_k is the number of balls of color k. Assume that the N balls are separated into two smaller boxes, so that box 1 has n balls and box 2 has the remaining $N-n$ balls. The statistician can observe the composition of box 1, represented by vector x of sample frequencies. The quantity of interest for the statistician is the vector $\psi - x$ representing the composition of box 2.

As in the bivariate case, we assume that $U^{1:N}$ is a finite sub-sequence in an exchangeable process and, therefore, any sub-sequence extracted from $U^{1:N}$ has the same distribution of $U^{1:n}$. Hence, $x = U^{1:n}\mathbf{1}$ has the same distribution of the frequency vector for a sample of size n.

As in the bivariate case, our objective is to find the distribution of $x|\psi$. Again, using de Finetti theorem, there is a vector $\mathbf{0} \le \theta \le \mathbf{1},\ \mathbf{1}'\theta = 1,$ such that $\coprod_{j=0}^{N} u^j \,|\, \theta$ and $\Pr\big(c(u^j) = k\big) = \theta_k$.

Theorem 4.1. *As in the Multinomial case, the following results follow:*

- $\psi|\theta \sim \mathrm{Mn}_m(N, \theta)$;
- $x|\theta \sim \mathrm{Mn}_m(n, \theta)$;
- $(\psi - x)\,|\,\theta \sim \mathrm{Mn}_m\big((N-n), \theta\big)$;
- $(\psi - x)$ II $x|\theta$.

Using the results of the last section and following the same steps as in the Hy_2 case in the first section, we obtain the following expression for m-variate Hypergeometric distribution, $x^n \,|\, [n, N, \psi] \sim \mathrm{Hy}_m(n, N, \psi)$:

$$\Pr\big(x^n | n, \psi\big) = \frac{\binom{n}{x^n}\binom{N-n}{\psi - x^n}}{\binom{N}{\psi}}$$

$$\text{where} \quad \mathbf{0} \le x^n \le \psi \le N\mathbf{1}, \quad \mathbf{1}'x^n = n, \quad \mathbf{1}'\psi = N\,.$$

This is the vector representation of the Hypergeometric probability distribution:

$$x^n \,|\, [n, x^N] \sim \mathrm{Hy}(n, N, x^N)\ .$$

Alternatively, we can write the more usual formula,

$$\Pr(x \,|\, \psi) = \frac{\binom{\psi_1}{x_1}\binom{\psi_2}{x_2}\cdots\binom{\psi_m}{x_m}}{\binom{N}{n}}\ .$$

Theorem 4.2. *The expectation and covariance of a random vector with Hypergeometric distribution, $x \sim \mathrm{Hy}_m(n, N, \psi)$, are:*

$$\mathrm{E}(x) = n\widetilde{\psi}\,, \qquad \mathrm{Cov}(x) = n\,\frac{N-n}{N-1}\left(\mathrm{diag}(\widetilde{\psi}) - \widetilde{\psi}\otimes\widetilde{\psi}'\right) \qquad \textit{where} \quad \widetilde{\psi} = \frac{1}{N}\,\psi\ .$$

Proof: Use that

$$\begin{aligned}
\mathrm{Cov}(x^n) &= n\,\mathrm{Cov}(u^1) + n(n-1)\,\mathrm{Cov}(u^1, u^2)\ , \\
\mathrm{Cov}(u^1) &= \mathrm{E}\big(u^1\otimes(u^1)'\big) - \mathrm{E}(u^1)\otimes\mathrm{E}(u^1)' = \mathrm{diag}(\widetilde{\psi}) - \widetilde{\psi}\otimes\widetilde{\psi}' \\
\mathrm{Cov}(u^1, u^2) &= \mathrm{E}\big(u^1\otimes(u^2)'\big) - \mathrm{E}(u^1)\otimes\mathrm{E}(u^2)'\ .
\end{aligned}$$

The second term of the last two equations are equal, and the first term of the last equation is

$$\mathrm{E}(u_i^1\,u_j^2) = \begin{cases} \dfrac{\psi_i}{N}\,\dfrac{\psi_i - 1}{N-1} & \text{if} \quad i = j\,, \\[2ex] \dfrac{\psi_i}{N}\,\dfrac{\psi_j}{N-1} & \text{if} \quad i \neq j\ . \end{cases}$$

Algebraic manipulation yields the result. □

Note that, as in the order 2 case, the diagonal elements of $\mathrm{Cov}(u^1)$ are positive, while the diagonal elements of $\mathrm{Cov}(u^1, u^2)$ are negative. In the off diagonal elements, the signs are reversed.

5. DIRICHLET DISTRIBUTION

In the second section we presented the multinomial distribution, $\mathrm{Mn}_m(n, \theta)$. In this section we present the Dirichlet distribution for the parameter θ. Let us first recall the univariate Poisson and Gamma distributions.

A random variable has Gamma distribution, $x\,|\,[a, b] \sim G(a, b),\ a, b > 0$, if its distribution is continuous with density

$$f(x\,|\,a, b) = \frac{b^a}{\Gamma(a)}\,x^{a-1}\exp(-bx)\ , \qquad x > 0\ .$$

The expectation and variance of this variate are

$$E(x) = \frac{a}{b} \qquad \text{and} \qquad \mathrm{Var}(x) = \frac{a}{b^2}\ .$$

Lemma 5.1. *Reproductive property for the Gamma distribution.*
If n independent random variables $x_i\,|\,a_i, b \sim G(a_i, b)$, then

$$\mathbf{1}'x \ \sim\ G(\mathbf{1}'a, b)\ .$$

Lemma 5.2. *The Gamma distribution is conjugate to the Poisson distribution.*

Proof: If $y|\lambda \sim \text{Ps}(\lambda)$ and λ has prior $\lambda|a,b \sim G(a,b)$, then

$$f(\lambda|y,a,b) \;\propto\; L(\lambda|y)\, f(\lambda) \;=$$
$$=\; \exp(-\lambda)\frac{\lambda^y}{y!}\;\frac{b^a}{\Gamma(a)}\,\lambda^{a-1}\exp(-b\lambda) \;\propto\; \lambda^{y+a-1}\exp\big(-(b+1)\,\lambda\big)\;. \quad \square$$

That is, the posterior distribution of λ is Gamma with parameters $[a+y,\, b+1]$.

Definition 5.1. Dirichlet distribution.
A random vector

$$y \in \mathcal{S}_{m-1} \;\equiv\; \Big\{ y \in \mathbb{R}^m \,|\; \mathbf{0} \le y \le \mathbf{1} \;\wedge\; \mathbf{1}'y = 1 \Big\}$$

has Dirichlet distribution of order m with positive $a \in \mathbb{R}^m$ if its density is

$$\Pr(y|a) \;=\; \frac{y \triangle (a-\mathbf{1})}{B(a)}\;.$$

Note that $\mathcal{S}_{m-1}$, the $m-1$ dimensional Simplex, is the region of $\mathbb{R}^m$ subject to the "constraint", $\mathbf{1}'y = 1$. Hence, a point in the Simplex has only $m-1$ "degrees of freedom". In this sense we say that the Dirichlet distribution has a "singular" representation. It is possible to give a non-singular representation to the distribution $[y_1, ..., y_{m-1}]'$, known as the Multivariate Beta distribution, but at the cost of obtaining a convoluted algebraic formulation that also loses the natural geometric interpretation of the singular form.

The normalization factor for the Dirichlet distribution is

$$B(a) \;\equiv\; \int_{y \in \mathcal{S}_{m-1}} \Big(y \triangle (a-1) \Big)\, dy\;.$$

Lemma 5.3. *Beta function.*
The normalization factor for the Dirichlet distribution defined above is the Beta function, defined as

$$B(a) \;=\; \frac{\prod_{k=1}^m \Gamma(a_k)}{\Gamma(\mathbf{1}'a)}\;.$$

The proof is given at the end of this section.

Theorem 5.1. *Dirichlet as Conjugate of the Multinomial.*
If $\theta \sim \mathrm{Di}_m(a)$ *and* $x|\theta \sim \mathrm{Mn}_m(n,\theta)$ *then*

$$\theta|x \sim \mathrm{Di}_m(a+x) \ .$$

Proof: We only have to remember that the Multinomial likelihood is proportional to $\theta \triangle x$, and that a Dirichlet prior is proportional to $\theta \triangle (a-\mathbf{1})$. Hence, the posterior is proportional to $\theta \triangle (x+a-1)$. At the other hand, $B(a+x)$ is the normalization factor, i.e., equal to the integral on θ of $\theta \triangle (x+a-1)$, and so we have a Dirichlet density function, as defined above. □

Theorem 5.2. *Dirichlet Moments.*
If $\theta \sim \mathrm{Di}_m(a)$ *and* $p \in \mathbb{N}^m$, *then*

$$\mathrm{E}(\theta \triangle p) = \frac{B(a+p)}{B(a)} \ .$$

Proof:

$$\begin{aligned}\int_\Theta (\theta \triangle p)\, f(\theta|a)\, d\theta &= \frac{1}{B(a)} \int_\Theta (\theta \triangle p) \Big(\theta \triangle (a-1)\Big)\, d\theta \\ &= \frac{1}{B(a)} \int_\Theta \Big(\theta \triangle (a+p-1)\Big)\, d\theta = \frac{B(a+p)}{B(a)} \ . \end{aligned}$$ □

Choosing the exponents, p, appropriately, we have

Corollary 5.1. *If* $\theta \sim \mathrm{Di}_m(a)$, *then*

$$\begin{aligned}\mathrm{E}(\theta) &= \widetilde{a} \equiv \frac{1}{\mathbf{1}'a}\, a \ , \\ \mathrm{Cov}(\theta) &= \frac{1}{\mathbf{1}'a+1} \Big(\mathrm{diag}(\widetilde{a}) - \widetilde{a} \otimes \widetilde{a}'\Big) \ . \end{aligned}$$

Theorem 5.3. *Characterization of the Dirichlet by the Gamma.*
Let the components of the random vector $x \in \mathbb{R}^m$ *be independent variables with distribution* $G(a_k, b)$. *Then, the normalized vector*

$$y = \frac{1}{\mathbf{1}'x}\, x \sim \mathrm{Di}_m(a) \ , \qquad \mathbf{1}'x \sim \mathrm{Ga}(\mathbf{1}'a) \qquad \textit{and} \qquad y \amalg \mathbf{1}'x \ .$$

Proof: Consider the normalization

$$y = \frac{1}{t}\, x \ , \qquad t = \mathbf{1}'x \ , \qquad x = t\, y \ ,$$

as a transformation of variables. Note that one of the new variables, say $y_m \equiv t(1 - y_1 \cdots - y_{m-1})$, becomes redundant.

The Jacobian matrix of this transformation is

$$J = \frac{\partial\,(x_1, x_2, ..., x_{m-1}, x_m)}{\partial\,(y_1, y_2, ..., y_{m-1}, t)} = \begin{bmatrix} t & 0 & \cdots & 0 & y_1 \\ 0 & t & \cdots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & t & y_{m-1} \\ -t & -t & \cdots & -t & 1 - y_1 \cdots - y_{m-1} \end{bmatrix} .$$

By elementary operations that add all rows to the last one, we obtain the LU factorization of the Jacobian matrix, $J = LU$, where

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & \cdots & -1 & 1 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} t & 0 & \cdots & 0 & y_1 \\ 0 & t & \cdots & 0 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & t & y_{m-1} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} .$$

A triangular matrix determinant is equal to the product of the elements in its main diagonal, hence $|J| = |L|\,|U| = 1\,t^{m-1}$.

At the other hand, the joint distribution of x is

$$f(x) = \prod_{k=1}^{m} \mathrm{Ga}(x_k|a_k, b) = \prod_{k=1}^{m} \frac{b^{a_k}}{\Gamma(a_k)}\, e^{-bx_k} (x_k)^{a_k - 1}$$

and the joint distribution in the new system of coordinates is

$$\begin{aligned} g([y,t]) &= |J|\, f\big(x^{-1}([y,t])\big) \\ &= t^{m-1} \prod_{k=1}^{m} \frac{b^{a_k}}{\Gamma(a_k)}\, e^{-bx_k} (x_k)^{a_k - 1} = t^{m-1} \prod_{k=1}^{m} \frac{b^{a_k}}{\Gamma(a_k)}\, e^{-bty_k} (ty_k)^{a_k - 1} \\ &= \left(\prod_{k=1}^{m} \frac{(y_k)^{a_k - 1}}{\Gamma(a_k)} \right) b^{\mathbf{1}'a}\, e^{-bt}\, t^{\mathbf{1}'a - m}\, t^{m-1} = \left(\prod_{k=1}^{m} \frac{(y_k)^{a_k - 1}}{\Gamma(a_k)} \right) b^{\mathbf{1}'a}\, e^{-bt}\, t^{\mathbf{1}'a - 1} . \end{aligned}$$

Hence, the marginal distribution of $y = [y_1, ..., y_k]'$ is

$$\begin{aligned} g(y) &= \int_{t=0}^{\infty} g([y,t])\, dt \\ &= \left(\prod_{k=1}^{m} \frac{(y_k)^{a_k - 1}}{\Gamma(a_k)} \right) \int_{t=0}^{\infty} b^{\mathbf{1}'a}\, e^{-bt}\, t^{\mathbf{1}'a - 1}\, dt \\ &= \left(\prod_{k=1}^{m} \frac{(y_k)^{a_k - 1}}{\Gamma(a_k)} \right) \Gamma(\mathbf{1}'a) = \frac{y \triangle (a-1)}{B(a)} . \end{aligned}$$

In the last passage, we have replaced the integral by the normalization factor of a Gamma density, $\mathrm{Ga}(\mathbf{1}'a, b)$. Hence, we obtain a density proportional to $y \triangle (a-1)$, i.e., a Dirichlet. □

In the last passage we also obtain the Dirichlet normalization factor, proving the Beta function lemma.

Lemma 5.4. *Bipartition of Indices for the Dirichlet.*
Let $1{:}t$, $t{+}1{:}m$ be a bipartition of the class index domain, $1{:}m$, of an order m Dirichlet, in two super-classes. Let $y \sim \mathrm{Di}_m(a)$, and

$$z^1 = \frac{1}{\mathbf{1}' y_{1:t}}\, y_{1:t} \ , \qquad z^2 = \frac{1}{\mathbf{1}' y_{t+1:m}}\, y_{t+1:m} \ , \qquad w = \begin{bmatrix} \mathbf{1}' y_{1:t} \\ \mathbf{1}' y_{t+1:m} \end{bmatrix} .$$

We then have $z^1 \amalg z^2 \amalg w$ and

$$z^1 \sim \mathrm{Di}_t(a_{1:t}) \ , \qquad z^2 \sim \mathrm{Di}_{m-t}(a_{t+1:m}) \qquad \text{and} \qquad w \sim \mathrm{Di}_2\left(\begin{bmatrix} \mathbf{1}' a_{1:t} \\ \mathbf{1}' a_{t+1:m} \end{bmatrix} \right) .$$

Proof: From the Dirichlet characterization by the Gamma we can imagine that the vector y is built by normalizing of a vector x, as follows:

$$y = \frac{1}{\mathbf{1}' x}\, x \ , \qquad x_k \sim \mathrm{Ga}(a_k, b) \ , \qquad \coprod_{k=1}^{m} x_k \ .$$

Considering separately each one of the super-classes, we build the vectors z^1 and z^2 that are distributed as

$$\begin{aligned} z^1 \ &= \ \frac{1}{\mathbf{1}' y_{1:t}}\, y_{1:t} \ = \ \frac{1}{\mathbf{1}' x_{1:t}}\, x_{1:t} \ \sim \ \mathrm{Di}_t(a_{1:t}) \ , \\ z^2 \ &= \ \frac{1}{\mathbf{1}' y_{t+1:m}}\, y_{t+1:m} \ = \ \frac{1}{\mathbf{1}' x_{t+1:m}}\, x_{t+1:m} \ \sim \ \mathrm{Di}_{m-t}(a_{t+1:m}) \ . \end{aligned}$$

$z^1 \amalg z^2$, that are in turn independent of the partial sums

$$\mathbf{1}' x_{1:t} \ \sim \ \mathrm{Ga}(\mathbf{1}' a_{1:t}, b) \qquad \text{and} \qquad \mathbf{1}' x_{t+1:m} \ \sim \ \mathrm{Ga}(\mathbf{1}' a_{t+1:m}, b) \ .$$

Using again the theorem characterizing the Dirichlet by the Gamma distribution for these two Gamma variates, we obtain the result. □

We can generalize this result for any partition of the set of classes, as follows. If $y \sim \mathrm{Di}_m(a)$ and T is a s-partition of the m classes, the intra and extra super-class distributions are independent Dirichlets, as follows:

$$\begin{aligned} z^r \ &= \ \frac{1}{T_r y}\, {}_rPy \ \sim \ \mathrm{Di}_{T_r 1}({}_rPa) \ , \\ w \ &= \ Ty \ \sim \ \mathrm{Di}_s(Ta) \ . \end{aligned}$$

6. DIRICHLET-MULTINOMIAL

We say that a random vector $x \in \mathbb{N}^n \,|\, \mathbf{1}'x = n$ has Dirichlet-Multinomial (DM) distribution with parameters n and $a \in \mathbb{R}^m$, iff

$$\Pr(x\,|\,n,a) \;=\; \frac{B(a+x)}{B(a)} \binom{n}{x} \;=\; \frac{B(a+x)}{B(a)\,B(x)} \,\frac{1}{x \triangle \mathbf{1}} \;.$$

Theorem 6.1. *Characterization of the DM as a Dirichlet mixture of Multinomials.*

If $\theta \sim \mathrm{Di}_m(a)$ *and* $x\,|\,\theta \sim \mathrm{Mn}(n,\theta)$ *then* $x \,|\, [n,a] \sim \mathrm{DM}_m(n,a)$.

Proof: The joint distribution of θ, x is proportional to $\theta \triangle (a+x-1)$, which integrated on θ is $B(a+x)$. Hence, multiplying by the joint distribution constants, we have the marginal for x, Q.E.D. Therefore, we have also proved that the function DM is normalized, that is

$$\begin{aligned}\Pr(x) \;&=\; \int_{\theta\in\mathcal{S}_{m-1}} \binom{n}{x} \,(\theta \triangle x)\, \frac{1}{B(a)}\, \theta \triangle (a-\mathbf{1})\; d\theta \\ &=\; \frac{1}{B(a)} \binom{n}{x} \int_{\theta\in\mathcal{S}_{m-1}} \Big(\theta \triangle (x+a-\mathbf{1})\Big)\, d\theta \;=\; \frac{B(x+a)}{B(a)} \binom{n}{x} \;. \qquad \square\end{aligned}$$

Theorem 6.2. *Characterization of the DM by m Negative Binomials.*
Let $a \in \mathbb{N}_+^m$, and $x \in \mathbb{N}_m$, be a vector whose components are independent random variables, $a_k \sim \mathrm{NB}(a_k,\theta)$. Then

$$x \,|\, [\mathbf{1}'x = n, a] \;\sim\; \mathrm{DM}_m(n,a) \;.$$

Proof:

$$\begin{aligned}\Pr(x\,|\,\theta,a) \;&=\; \prod_{k=1}^{m} \binom{a_k + x_k - 1}{x_k}\, \theta^{a_k}(1-\theta)^{x_k} \;, \\ \Pr(\mathbf{1}'x\,|\,\theta,a) \;&=\; \binom{\mathbf{1}'a + \mathbf{1}'x - 1}{\mathbf{1}'x}\, \theta^{\mathbf{1}'a}(1-\theta)^{\mathbf{1}'a} \;.\end{aligned}$$

Then,

$$\Pr\big(x \,|\, \mathbf{1}'x = n, \theta, a\big) \;=\; \frac{\Pr(x\,|\,a,\theta)}{\Pr(\mathbf{1}'x = n \,|\, \theta)} \;=\; \frac{\prod_{k=1}^{m} \binom{a_k + x_k - 1}{x_k}}{\binom{\mathbf{1}'a + \mathbf{1}'x - 1}{\mathbf{1}'x}} \;.$$

Hence,

$$\begin{aligned}\Pr\big(x \,|\, \mathbf{1}'x = n,\, \theta,\, a\big) \;&=\; \Pr\big(x \,|\, \mathbf{1}'x = n,\, a\big) \\ &= \prod_{k=1}^{m} \frac{\Gamma(a_k + x_k)}{x!\;\Gamma(a_k)} \Big/ \frac{\Gamma(\mathbf{1}'a + n)}{\Gamma(\mathbf{1}'a)\;n!} \;=\; \frac{B(a+x)}{B(a)} \binom{n}{x} \,. \qquad \square\end{aligned}$$

Theorem 6.3. *The DM as Pseudo-Conjugate for the Hypergeometric.*

If $x \sim \mathrm{Hy}_m(n, N, \psi)$ *and* $\psi \sim \mathrm{DM}_m(N, a)$ *then* $(\psi - x) \,|\, x \sim \mathrm{DM}_m(N - n, a)$.

Proof: Using the properties of the Hypergeometric already presented, we have the independence relation, $(\psi - x) \amalg x|\theta$. We can therefore use the Multinomial sample $x|\theta$ for updating the prior and obtain the posterior

$$\theta|x \;\sim\; \mathrm{Di}_m(a + x) \,.$$

Hence, the distribution of the non sampled pat of the population, $\psi - x$, given the sample x, is a mixture of $(\psi - x)\theta$ by the posterior for θ. By the characterization of the DM as a mixture of Multinomials by a Dirichlet, the theorem follows, i.e.,

$$\left.\begin{aligned}(\psi - x) \,|\, [\theta, x] \;\sim\; (\psi - x) \,|\, \theta \;\sim\; \mathrm{Mn}_m(N - n, \theta) \\ \theta|x \;\sim\; \mathrm{Di}_m(a + x)\end{aligned}\right\} \;\Longrightarrow$$

$$\Longrightarrow \quad (\psi - x) \,|\, x \;\sim\; \mathrm{Di}_m(N - n, a + x) \,. \qquad \square$$

Lemma 6.1. *DM Expectation and Covariance.*
If $x \sim \mathrm{DM}_m(n, a)$ *then*

$$\begin{aligned}\mathrm{E}(x) \;&=\; n\,\widetilde{a} \;\equiv\; \frac{1}{\mathbf{1}'a}\,a \,, \\ \mathrm{Cov}(x) \;&=\; \frac{n(n + \mathbf{1}'a)}{\mathbf{1}'a + 1}\,\big(\mathrm{diag}(\widetilde{a}) - \widetilde{a} \otimes \widetilde{a}'\big) \,.\end{aligned}$$

Proof:

$$\mathrm{E}(x) \;=\; \mathrm{E}_\theta\big(\mathrm{E}_x(x|\theta)\big) \;=\; \mathrm{E}_\theta(n\theta) \;=\; n\widetilde{a} \,;$$

$$\begin{aligned}\mathrm{E}(x \otimes x') \;&=\; \mathrm{E}_\theta\big(\mathrm{E}_x(x \otimes x' \,|\, \theta)\big) \\ &=\; \mathrm{E}_\theta\Big(\mathrm{E}(x|\theta) \otimes \mathrm{E}(x|\theta)' + \mathrm{Cov}(x|\theta)\Big) \\ &=\; \mathrm{E}_\theta\Big(n\big(\mathrm{diag}(\theta) - \theta \otimes \theta'\big) + n^2\,\theta \otimes \theta'\Big) \\ &=\; n\,\mathrm{E}_\theta\big(\mathrm{diag}(\theta)\big) + n\,(n-1)\,\mathrm{E}_\theta(\theta \otimes \theta') \\ &=\; n\,\mathrm{diag}(\widetilde{a}) + n\,(n-1)\Big(\mathrm{E}(\theta) \otimes \mathrm{E}(\theta)' + \mathrm{Cov}(\theta)\Big) \\ &=\; n\,\mathrm{diag}(\widetilde{a}) + n\,(n-1)\left(\widetilde{a} \otimes \widetilde{a}' + \frac{1}{\mathbf{1}'a + 1}\,\big(\mathrm{diag}(\widetilde{a}) - \widetilde{a} \otimes \widetilde{a}'\big)\right) \\ &=\; n\,\mathrm{diag}(\widetilde{a}) + n(n-1)\left(\frac{1}{\mathbf{1}'a + 1}\,\mathrm{diag}(\widetilde{a}) + \frac{\mathbf{1}'a}{\mathbf{1}'a + 1}\;\widetilde{a} \otimes \widetilde{a}'\right) \,;\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(x) &= \mathrm{E}(x\otimes x') - \mathrm{E}(x)\otimes\mathrm{E}(x)' = \mathrm{E}(x\otimes x') - n^2\,\widetilde{a}\otimes\widetilde{a}' \\ &= \left(n + \frac{n(n-1)}{\mathbf{1}'a+1}\right)\mathrm{diag}(\widetilde{a}) + \left(n(n-1)\,\frac{\mathbf{1}'a}{\mathbf{1}'a+1} - n^2\right)\widetilde{a}\otimes\widetilde{a}' \\ &= \frac{n(n+\mathbf{1}'a)}{\mathbf{1}'a+1}\left(\mathrm{diag}(\widetilde{a}) - \widetilde{a}\otimes\widetilde{a}'\right) . \qquad \square\end{aligned}$$

Theorem 6.4. *DM Class Bipartition.*

Let $1{:}t$, $t{+}1{:}m$ a bipartition of the index domain for the classes of an order m DM, $1{:}m$, in two super-classes. Then, the following conditions (i) to (iii) are equivalent to condition (iv):

(i) $x_{1:t} \amalg x_{t+1:m} \,|\, n_1 = \mathbf{1}'x_{1:t}$;

(ii-1) $x_{1:t} \,|\, n_1 = \mathbf{1}'x_{1;t} \sim \mathrm{DM}_t(n_1, a_{1:t})$;

(ii-2) $x_{t+1:m} \,|\, n_2 = \mathbf{1}'x_{t+1:m} \sim \mathrm{DM}_{m-t}(n_2, a_{t+1:m})$;

(iii) $\begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \sim \mathrm{DM}_2\left(n, \begin{bmatrix} \mathbf{1}'a_{1:t} \\ \mathbf{1}'a_{t+1:m} \end{bmatrix}\right)$;

(iv) $x \sim \mathrm{DM}_m(n, a)$.

Proof: We only have to show that the joint distribution can be factored in this form. By the DM characterization as a mixture, we can write it as Dirichlet mixture of Multinomials. By the bipartition theorems, we can factor both, the Multinomials and the Dirichlet, so the theorem follows. $\square$

7. DIRICHLET OF THE SECOND KIND

Consider $y \sim \mathrm{Di}_{m+1}(a)$. The vector $z = (1/y_{m+1})y_{1:m}$ has Dirichlet of the Second Kind (D2K) distribution.

Theorem 7.1. *Characterization of D2K by the Gamma distribution.*

Using the characterization of the Dirichlet by the Gamma, we can write the D2K variate as a function of $m+1$ independent Gamma variates,

$$z_{1:m} \sim (1/x_{m+1})x_{1:m} \qquad \textit{where} \quad x_k \sim Ga(a_k, b) .$$

Similar to what we did for the Dirichlet (of the first kind), we can write the D2K distribution and its moments as:

$$\begin{aligned} f(z|a) &= \frac{z \triangle (a_{1:m} - 1)}{(1+\mathbf{1}'z)^{\mathbf{1}'a}\; B(a)} , \\ E(z) &= e = (1/a_{m+1})\, a_{1:m} , \\ \mathrm{Cov}(z) &= \frac{1}{a_{m+1}-2}\left(\mathrm{diag}(e) + e\otimes e'\right) . \end{aligned}$$

The logarithm of a Gamma variate is well approximated by a Normal variate, see Aitchison and Shen (1980). This approximation is the key to several efficient computational procedures, and motivates the computation of the first two moments of the log-D2K distribution. For that, we use the Digamma, $\psi(\ )$, and Trigamma function, $\psi'(\ )$, defined as:

$$\psi(a) = \frac{d}{da} \ln \Gamma(a) = \frac{\Gamma'(a)}{\Gamma(a)} \ , \qquad \psi'(a) = \frac{d}{da} \psi(a) \ .$$

Lemma 7.1. *The expectation and covariance of a log-D2K variate are:*

$$\begin{aligned} E\big(\log(z)\big) \ &= \ \psi(a_{1:m}) - \psi(a_{m+1})\, \mathbf{1} \ , \\ \mathrm{Cov}\big(\log(z)\big) \ &= \ \mathrm{diag}\big(\psi'(a_{1:m}) + \psi'(a_{m+1})\big)\, \mathbf{1} \otimes \mathbf{1}' \ . \end{aligned}$$

Proof: Consider a Gamma variate, $x \sim G(a, 1)$:

$$1 \ = \int_0^\infty f(x)\, dx \ = \int_0^\infty \frac{1}{\Gamma(a)}\, x^{a-1} \exp(-x)\ dx \ .$$

Taking the derivative with respect to parameter a, we have

$$0 \ = \int_0^\infty \ln(x)\, x^{a-1}\, \frac{\exp(-x)}{\Gamma(a)}\ dx \ - \ \frac{\Gamma'(a)}{\Gamma^2(a)}\, \Gamma(a) \ = \ E\big(\ln(x)\big) - \psi(a) \ .$$

Taking the derivative with respect to parameter a a second time,

$$\begin{aligned} \psi'(a) \ &= \ \frac{d}{da} E\big(\ln(x)\big) \ = \ \frac{d}{da} \int_0^\infty \frac{\ln(x)}{\Gamma(a)}\, x^{a-1} \exp(-x)\ dx \\ &= \ \int_0^\infty \ln(x)^2\, x^{a-1}\, \frac{\exp(-x)}{\Gamma(a)}\ dx \ - \ \frac{\Gamma'(a)}{\Gamma(a)}\, E\big(\ln(x)\big) \\ &= \ E\big(\ln(x)^2\big) - E\big(\ln(x)\big)^2 \ = \ \mathrm{Var}\big(\ln(x)\big) \ . \end{aligned}$$

The lemma follows from the D2K characterization by the Gamma. □

8. EXAMPLES

Example 8.1. Let A, B be two attributes, each one of them present or absent in the elements of a population. Then each element of this population can be classified in exactly one of $2^2 = 4$ categories:

A	B	k	I^k
present	present	1	$[1,0,0,0]'$
present	absent	2	$[0,1,0,0]'$
absent	present	3	$[0,0,1,0]'$
absent	absent	4	$[0,0,0,1]'$

According to the notation above, we can write $x|n,\theta \sim \text{Mn}_4(n,\theta)$.

If $\theta = [0.35,\ 0.20,\ 0.30,\ 0.15]$ and $n = 10$, then

$$\Pr\big(x^{10}|n,\theta\big) \;=\; \binom{10}{x^{10}}\,(\theta \triangle x^{10})\ .$$

Hence, in order to compute the probability of $x = [1,2,3,4]'$ given θ, we use the expression above, obtaining

$$\Pr\left(\begin{bmatrix}1\\2\\3\\4\end{bmatrix}\ \middle|\ \begin{bmatrix}0.35\\0.20\\0.30\\0.15\end{bmatrix}\right) \;=\; 0.000888\ .$$

Example 8.2. If $X|\theta \sim \text{Mn}_3(10,\theta)$, $\theta = [0.20,\ 0.30,\ 0.15]$, one can conclude, using the result above, that

$$\text{E}(X) \;=\; (2,3,1.5)\ ,$$

while the covariance matrix is

$$\Sigma \;=\; \begin{bmatrix} 1.6 & -0.6 & -0.3 \\ -0.6 & 2.1 & -0.45 \\ -0.3 & -0.45 & 1.28 \end{bmatrix}\ .$$

Example 8.3. Assume that $X|\theta \sim \text{Mn}_3(10,\theta)$, with $\theta = [0.20,\ 0.30,\ 0.15]$, as in Example 2. Let us take $A_0 = \{0,1\}$, $A_1 = \{2,3\}$. Then,

$$\sum_{A_1} X_i|\theta \;=\; X_2 + X_3\,|\,\theta \;\sim\; \text{Mn}_1(10,\theta_2+\theta_3)\ ,$$

or

$$X_2 + X_3\,|\,\theta \;\sim\; \text{Mn}_1(10,0.45)\ .$$

Analogously,

$$\begin{aligned} X_0 + X_1\,|\,\theta \;&\sim\; \text{Mn}_1(10,0.55)\ ,\\ X_1 + X_3\,|\,\theta \;&\sim\; \text{Mn}_1(10,0.35)\ ,\\ X_2\,|\,\theta \;&\sim\; \text{Mn}_1(10,0.30)\ . \end{aligned}$$

Note that, in general, if $X|\theta \sim \text{Mn}_k(n,\theta)$, then $X_i\,|\,\theta \sim \text{Mn}_1(n,\theta_i)$, for $i = 1,...,k$.

Example 8.4. 3×3 Contingency Tables.
Assume that $X\,|\,\theta \sim \text{Mn}_8(n,\theta)$, as in a 3×3 Contingency Tables:

x_{11}	x_{12}	x_{13}	$x_{1\bullet}$
x_{21}	x_{22}	x_{23}	$x_{2\bullet}$
x_{31}	x_{32}	x_{33}	$x_{3\bullet}$
$x_{\bullet 1}$	$x_{\bullet 2}$	$x_{\bullet 3}$	n

Applying Theorem 3.2 we get

$$(X_{1\bullet}, X_{2\bullet}) \,|\, \theta \;\sim\; \mathrm{Mn}_2(n, \theta') \;, \qquad \theta' = (\theta_{1\bullet}, \theta_{2\bullet}), \quad \theta'_0 = \theta_3 \;.$$

This result tell us that

$$(X_{i1}, X_{i2}, X_{i3}) \,|\, \theta \;\sim\; \mathrm{Mn}_3(n, \theta'_i) \;,$$

with

$$\theta'_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}), \quad \theta'_{0i} = 1 - \theta_{i\bullet} \;, \qquad i = 1, 2, 3 \;.$$

We can now apply Theorem 3.3 to obtain the probability distribution of each row of the contingency table, conditioned on its sum, or conditioned on the sum of the other rows. We have

$$(X_{i1}, X_{i2}) \,|\, x_{i\bullet} \;, \qquad \theta \sim \mathrm{Mn}_2(x_{i\bullet}, \theta'_i)$$

with

$$\theta'_i = \frac{(\theta_{il}, \theta_{i2})}{\theta_{i\bullet}} \;, \qquad \theta'_{0i} = \frac{\theta_{i3}}{\theta_{i\bullet}} \;.$$

The next result expresses the distribution of $X \,|\, \theta$ in term of the conditional distributions, of each row of the table, in its sum, and in term of the distribution of these sums.

Proposition 8.1. *If $X \,|\, \theta \sim \mathrm{Mn}_{r^2-1}(n, \theta)$, as in an $r \times r$, contingency table, then $P(X \,|\, \theta)$ can be written as*

$$P(X \,|\, \theta) \;=\; \left[\prod_{i=1}^{r} P\big(X_{i1}, ..., X_{i,r-1} \,|\, x_{i\bullet}, \theta\big) \right] P\big(X_{1\bullet}, ..., X_{r-1\bullet} \,|\, \theta\big) \;.$$

Proof: We have:

$$\begin{aligned} P(X \,|\, \theta) \;&=\; n! \prod_{i=1}^{r} \frac{\theta_i^{x_i}}{x_i!} \;=\; n! \, \frac{\theta_{11}^{x_{11}} \cdots \theta_{rr}^{x_{rr}}}{x_{11}! \cdots x_{rr}!} \\ &=\; \left[\prod_{i=1}^{r} \frac{x_{i\bullet}!}{x_{i1}! \cdots x_{ir}!} \left(\frac{\theta_{i1}}{\theta_{i\bullet}}\right)^{x_{i1}} \cdots \left(\frac{\theta_{ir}}{\theta_{i\bullet}}\right)^{x_{ir}} \right] \frac{n!}{x_{i\bullet}! \cdots x_{r\bullet}!} \; \theta_{1\bullet}^{x_{1\bullet}} \cdots \theta_{r\bullet}^{x_{r\bullet}} \;. \end{aligned}$$

From Theorems 3.2 and 3.3, as in the last example, we recognize each of the first r factors above as the probabilities of each row in the table, conditioned on its sum, and recognize the last factor as the joint probability distribution of sum of these r rows. □

Corollary 8.1. *If* $X|\theta \sim \mathrm{Mn}_{r^2-1}(n,\theta)$, *as in Theorems 3.2 and 3.3, then*

$$P\big(X \,|\, x_{1\bullet}, ..., x_{r-1\bullet}\,,\, \theta\big) \;=\; \prod_{i=1}^{r} P\big(X_{i1}, ..., X_{i,r-1} \,|\, x_{i\bullet}\,,\, \theta\big)$$

and, knowing θ, $x_{1\bullet}, ..., x_{r-1\bullet}$,

$$(X_{11}, ..., X_{1,r-1}) \amalg ... \amalg (X_{r1}, ..., X_{r,r-1}) \;.$$

Proof: Since

$$P(X|\theta) \;=\; P\big(X \,|\, x_{1\bullet}, ..., x_{r-1\bullet}\,,\, \theta\big)\, P\big(X_{1\bullet}, X_{2\bullet}, ..., X_{r-1\bullet} \,|\, \theta\big) \;,$$

from Theorems 3.2 and 3.3 we get the proposed equality. □

The following result will be used next to express Theorem 3.4 as a canonical representation for $P(X|\theta)$.

Proposition 8.2. *If* $X|\theta \sim \mathrm{Mn}_{r^2-1}(n,\theta)$, *as in Proposition, then a transformation*

$$T:\; \big(\theta_{11}, ..., \theta_{1r}, ..., \theta_{r1}, ..., \theta_{r,r-1}\big) \;\to\; \big(\lambda_{11}, ..., \lambda_{1,r-1}, ..., \lambda_{r1}, ..., \lambda_{r,r-1}, \eta_1, ..., \eta_{r-1}\big)$$

given by

$$\begin{aligned} \lambda_{11} = \frac{\theta_{11}}{\theta_{1\bullet}} \;,\quad \cdots \;,\quad \lambda_{1,r-1} = \frac{\theta_{1,r-1}}{\theta_{1\bullet}} \\ \vdots \\ \lambda_{r1} = \frac{\theta_{r1}}{\theta_{r\bullet}} \;,\quad \cdots \;,\quad \lambda_{r,r-1} = \frac{\theta_{r,r-1}}{\theta_{r\bullet}} \end{aligned}$$

$$\eta_1 = \theta_{1\bullet}\,,\quad \eta_2 = \theta_{2\bullet}\,,\quad \ldots,\quad \eta_{r-1} = \theta_{(r-1)\bullet}$$

is a onto transformation defined in $\big\{0 < \theta_{11} + \cdots + \theta_{r,r-1} < 1\,;\;\; 0 < \theta_{ij} < 1\big\}$ *over the unitary cube of dimension* r^2-1. *Moreover, the Jacobian of this transformation, t, is*

$$J \;=\; \eta^{r-1}\, \eta_1^{r-1} \cdots\, \eta_{r-1}^{r-1}\, \big(1 - \eta_1 - \cdots - \eta_{r-1}\big)^{r-1} \;.$$

The proof is not hard to check.

Example 8.5. Let us examine the case of a 2×2 contingency table:

x_{11}	x_{12}
x_{21}	x_{22}

n

θ_{11}	θ_{12}
θ_{21}	θ_{22}

1

In order to obtain the canonical representation of $P(X|\theta)$ we use the transformation T in the case $r = 2$:

$$\lambda_{11} = \frac{\theta_{11}}{\theta_{11} + \theta_{12}} ,$$

$$\lambda_{21} = \frac{\theta_{11}}{\theta_{21} + \theta_{22}} ,$$

$$\eta_1 = \theta_{11} + \theta_{12} ,$$

hence,

$$P(X|\theta) = \binom{x_{1\bullet}}{x_{11}} \lambda_{11}^{x_{11}} (1-\lambda_{11})^{x_{12}} \binom{x_{2\bullet}}{x_{21}} \lambda_{21}^{x_{21}} (1-\lambda_{21})^{x_{22}} \binom{n}{x_{1\bullet}} \eta_1^{x_{1\bullet}} (1-\eta_1)^{x_{2\bullet}} ,$$

$$0 < \theta_{11} < 1, \quad 0 < \theta_{21} < 1, \quad 0 < \eta_1 < 1 .$$

9. FUNCTIONAL CHARACTERIZATIONS

The objective of this section is to derive the general form of a homogeneous Markov random process. Theorem 9.1, by Reny and Aczel, states that such a process is described by a mixture of Poisson distributions. Our presentation follows Aczél (1966, Sec. 2.1 and 2.3) and Jánossy, Rényi and Aczél (1950). It follows from the characterization of the Multinomial by the Poisson distribution given in Theorem 3.1, that Reny–Aczel characterization of a homogeneous and local time point process is analogous to de Finetti characterization of an infinite exchangeable 0-1 process as a mixture of Bernoulli distributions, see for example Feller (V. 2, Ch. VII, Sec. 4).

Cauchy's Functional Equations

Cauchy's additive functional equation has the form

$$f(x + y) = f(x) + f(y) .$$

The following argument from Cauchy (1821) shows that a continuous solution of this functional equation must have the form

$$f(x) = c\,x .$$

Repeating the sum of the same argument, x, n times, we must have $f(nx) = n\,f(x)$. If $x = (m/n)t$, then $nx = mt$ and

$$n\,f(x) = f(nx) = f(mt) = m\,f(t) ,$$

hence

$$f\left(\frac{m}{n}\,t\right) = \frac{m}{n}\,f(t)\ ,$$

taking $c = f(1)$, and $x = m/n$, it follows that $f(x) = cx$, over the rationals, $x \in \mathbb{Q}$. From the continuity condition for $f(x)$, the last result must also be valid over the reals, $x \in \mathbb{R}$. Q.E.D.

Cauchy's multiplicative functional equation has the form

$$f(x+y) = f(x)\,f(y)\ , \qquad \forall\, x, y > 0\,, \quad f(x) \geq 0\ .$$

The trivial solution of this equation is $f(x) \equiv 0$. Assuming $f(x) > 0$, we take the logarithm, reducing the multiplicative equation to the additive equation,

$$\ln f(x_y) = \ln f(x) + \ln f(y)\ ,$$

hence

$$\ln f(x) = cx\,, \qquad \text{or} \qquad f(x) = \exp(cx)\ .$$

Homogeneous Discrete Markov Processes

We seek the general form of a homogeneous discrete Markov process. Let $w_k(t)$, for $t \geq 0$, be the probability of occurrence of exactly k events. Let us also assume the following hypotheses:

Time Locality: If $t_1 \leq t_2 \leq t_3 \leq t_4$ then, the number of events in $[t_1, t_2[$ is independent of the number of events in $[t_3, t_4[$.

Time Homogeneity: The distribution for the number of events occurring in $[t_1, t_2[$ depends only on the interval length, $t = t_2 - t_1$.

From time locality and homogeneity, we can decompose the occurrence of no (zero) events in $[0, t+u[$ as ,

$$w_0(t+u) = w_0(t)\,w_0(u)\ .$$

Hence, $w_0(t)$ must obey Cauchy's functional equation, and

$$w_0(t) = \exp(ct) = \exp(-\lambda t)\ .$$

Since $w_0(t)$ is a probability distribution, $w_0(t) \leq 1$, and $\lambda > 0$.

Hence, $v(t) = 1 - w_0(t) = 1 - \exp(-\lambda t)$, the probability of one or more events occurring before $t > 0$, must be the familiar exponential distribution.

For $k \geq 1$ occurrences before $t + u$, the general decomposition relation is

$$w_n(t+u) = \sum_{k=0}^{n} w_k(t)\,w_{n-k}(u)\ .$$

Theorem 9.1 (Reny–Aczel)**.** *The general (non trivial) solution of this this system of functional equations has the form:*

$$w_k(t) \,=\, e^{-\lambda t} \sum_{\langle r,k\rangle} \prod_{j=1}^{k} \frac{(c_j\, t)^{r_j}}{r_j!} \;, \qquad \lambda = \sum_{j=1}^{\infty} c_j \;.$$

where the index set $\langle r, k, n\rangle$ *is defined as*

$$\langle r, k, n\rangle \,=\, \Big\{ r_1, r_2, ..., r_k \,\Big|\; r_1 + 2r_2 + \cdots + k\, r_k = n \Big\} \;.$$

and $\langle r, k\rangle$ *is a shorthand for* $\langle r, k, k\rangle$*.*

Proof: By induction: The theorem is true for $k = 0$. Let us assume, as induction hypothesis, that it is true to $k < n$. The last equation in the recursive system is

$$\begin{aligned} w_n(t+u) \;&=\; \sum_{k=0}^{n} w_k(t)\, w_{n-k}(u) \;= \\ &=\; w_n(t)\, e^{-\lambda u} + w_n(u)\, e^{-\lambda t} + e^{-\lambda(t+u)} \sum_{k=1}^{n-1} \sum_{\langle r,k\rangle} \sum_{\langle s,n-k\rangle} \prod_{i=1}^{k} \frac{(c_i\, t)^{r_i}}{r_i!} \prod_{j=1}^{k} \frac{(c_j\, u)^{s_j}}{s_j!} \;. \end{aligned}$$

Defining

$$f_n(t) \,=\, e^{\lambda t}\, w_n(t) - \sum_{\langle r,n-1,n\rangle} \prod_{j=1}^{n-1} \frac{(c_j\, t)^{r_j}}{r_j!} \;,$$

the recursive equation takes the form

$$f_n(t+u) \,=\, f_n(t) + f_n(u) \;,$$

and can be solved as a general Cauchy's equation, that is,

$$f_n(t) \,=\, c_n\, t \;.$$

From the last equation and the definition of $f_n(t)$, we get the expression of $w_n(t)$ as in Theorem 9.1. The constant λ is chosen so that the distribution is normalized. □

The general solution given by Theorem 9.1 represents a composition (mixture) of Poisson processes, where an event in the j-th process in the composition corresponds to the simultaneous occurrence of j single events in the original homogeneous Markov process. If we impose the following rarity condition, the general solution is reduced to a mixture of ordinary Poisson processes.

Rarity Condition: The probability that an event occurs in a short time at least once is approximately equal to the probability that it occurs exactly once, that is, the probability of simultaneous occurrences is zero.

10. FINAL REMARKS

This work is in memory of Professor D. Basu who was the supervisor of the first author PhD dissertation, the starting point for the research in Bayesian analysis of categorical data presented here. A long list of papers follows Basu and Pereira (1982). We have chosen a few that we recommend for additional reading: Albert (1985), Gunel (1984), Irony, Pereira and Tiwari (2000), Paulino and Pereira (1992, 1995) and Walker (1996). To make the analysis more realistic, extensions and mixtures of Dirichlet also were considered. For instance see Albert and Gupta (1983), Carlson (1977), Dickey (1983), Dickey, Jiang and Kadane (1987), and Jiang, Kadane and Dickey (1992).

Usually the more complex distributions are used to realistic represent situations for which the strong properties of Dirichlet seems to be not realistic. For instance, in a 2×2 contingency table, the first line to be conditional independent of the second line given the marginal seems to be unrealistic in some situations. Mixtures of Dirichlet in some cases take care of the situation as shown by Albert and Gupta (1983).

The properties presented here are also important in non-parametric Bayesian statistics in order to understand the Dirichlet process for the competitive risk survival problem. See for instance Salinas-Torres, Pereira and Tiwari (1997, 2002). In order to be historically correct we cannot forget the important book of Wilks, published in 1962, where one can find the definition of Dirichlet distribution.

This article adopts a singular notation and representation, first used in Pereira and Stern (2005). Singular representations are unusual in statistical texts. Nevertheless, the singular notation makes it simpler to extend and generalize theoretical results and greatly facilitates numerical and computational implementation.

We end this article presenting the Reny–Aczel characterization of the Poisson mixture. This result can be interpreted as an alternative to de Finetti characterization theorem introduced in Finetti (1937). Using the characterization of binomial distributions by Poisson processes conditional arguments, as given by Theorem 3.1, and Blackwell (minimal) sufficiency properties discussed in Basu and Pereira (1983), Section 9 leads in fact to a De Finetti characterization for Binomial distributions. Also, if one recall the indifference principle (Mendel, 1989) the finite version of Finetti argument can simply be obtained. See also Irony and Pereira (1994) for the motivation of these arguments. The consideration of Section 9 could be viewed as a very simple formulation of the binomial distribution finite characterization.

ACKNOWLEDGMENTS

The authors had support from IME-USP, FAPESP and CNPq.

REFERENCES

[1] Aczél, J. (1966). *Lectures on Functional Equations and their Applications*, New York, Academic Press.

[2] Albert, J.H. (1985). *Bayesian estimation methods for incomplete two-way contingency tables using prior belief of association*. In "Bayesian Statistics" (J.M. Bernardo, ... and A.F.M. Smith, Eds.), Amsterdam, North Holland, 2:589–602.

[3] Albert, J.H. and Gupta, A.K. (1983). Bayesian estimation methods for 2×2 contingency tables using mixtures of Dirichlet distributions, *JASA*, **78**, 831–841.

[4] Aitchison, J. (2003). *The Statistical Analysis for Compositional Data*, (2nd Ed.), Blackburn Press, Caldwell.

[5] Aitchison, J. and Shen, S.M. (1980). Logistic-normal distributions: some properties and uses, *Biometrika*, **67**, 261–272.

[6] Basu, D. and Pereira, C.A.B. (1982). On the Bayesian analysis of categorical data: the problem of nonresponse, *JSPI*, **6**, 345–362.

[7] Basu, D. and Pereira, C.A.B. (1983). A note on Blackwell sufficiency and a Shibinsky characterization of distributions, *Sankhya A*, **45**(1), 99–104.

[8] de Finetti, B. (1947). La prévision: des lois logiques, ses sourses subjectives, *Annalles de l'Institut Henri Poincaré*, **7**, 1–68. English translation: *Foresight: Its logical laws, its subjective sources*. In "Studies in Subjective Probability" (Kiburg and Smoker, Eds.) (1963), pp. 93–158, Wiley, New York.

[9] Dickey, J.M. (1983). Multiple hypergeometric functions: probabilistic interpretations and statistical uses, *JASA*, **78**, 628–637.

[10] Dickey, J.M.; Jiang, T.J. and Kadane, J.B. (1987). Bayesian methods for categorical data, *JASA*, **82**, 773–781.

[11] Feller, W. (1957). *An Introduction to Probability Theory and Its Applications* (2nd Ed.), Vol. 1, Wiley, New York.

[12] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications* (2nd Ed.), Vol. 2, Wiley, New York.

[13] Gunel, E. (1984). A Bayesian analysis of the multinomial model for a dichotomous response with non-respondents, *Communications in Statistics – Theory and Methods*, **13**, 737–751.

[14] Irony, T.Z. and Pereira, C.A.B. (1994). Motivation for the use of discrete distributions in quality assurance, *Test*, **3**(2), 181–193.

[15] Irony, T.Z.; Pereira, C.A.B. and Tiwari, R.C. (2000). Analysis of opinion swing: comparison of two correlated proportions, *The American Statistician*, **54**, 57–62.

[16] Jánossy, L.; Rényi, A. and Aczél, J. (1950). On composed Poisson distributions, *Acta Math. Hungarica*, **1**, 209–224.

[17] Jiang, T.J.; Kadane, J.B. and Dickey, J.M. (1992). Computation of Carlson's multiple hypergeometric function R for Bayesian applications, *Journal of Computational and Graphical Statistics*, **1**, 231–251.

[18] Kadane, J.B. (1985). Is victimization chronic? A Bayesian analysis of multinomial missing data, *Journal of Econometrics*, **29**, 47–67.

[19] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

[20] Martin, J.J. (1975). *Bayesian decision and problems and Markov Chains.*

[21] Mendel, M.B. (1989). *Development of Bayesian parametric theory with application in control*, PhD Thesis, MIT, Cambridge: MA.

[22] Paulino, C.D.M. and Pereira, C.A.B. (1992). Bayesian analysis of categorical data informatively censored, *Communications in Statistics – Theory and Methods*, **21**, 2689–2705.

[23] Paulino, C.D.M. and Pereira, C.A.B. (1995). Bayesian methods for categorical data under informative general censoring, *Biometrika*, **82**(2), 439–446.

[24] Pereira, C.A.B. and Stern, J.M. (2005). Inferência indutiva com dados discretos: uma visão genuinamente Bayesiana, *COMCA-2005*, Universidad de Antofagasta, Chile.

[25] Salinas-Torres, V.H.S.; Pereira, C.A.B. and Tiwari, R.C. (1997). Convergence of Dirichlet measures arising in context of Bayesian analysis of competing risks models, *J. Multivariate Analysis*, **62**(1), 24–35.

[26] Salinas-Torres, V.H.S.; Pereira, C.A.B. and Tiwari, R.C. (2002). Bayesian nonparametric estimation in a series system or a competing-risks model, *J. of Nonparametric Statistics*, **14**(4), 449–458.

[27] Smith, P.J. and Gunel, E. (1984). Practical Bayesian approaches to the analysis of 2×2 contingency table with incompletely categorized Data, *Communication of Statistics – Theory and Methods*, **13**, 1941–63.

[28] Stern, J.M. (2007). *Cognitive constructivism and the epistemic significance of sharp statistical hypotheses*, 2007 Summer Program, Institute of Mathematics and Statistics, University of São Paulo, Brazil.

[29] Tian, G.L.; Ng, K.W. and Geng, Z. (2003). Bayesian computation for contingency tables with incomplete cells-counts, *Statistica Sinica*, **13**, 189-206.

[30] Walker, S. (1986). A Bayesian maximum posteriori algorithm for categorical data under informative general censoring, *The Statistician*, **45**, 293–298.

[31] Wilks, S.S. (1962). *Mathematical Statistics*, Wiley, New York.

REVSTAT – Statistical Journal
Volume 6, Number 3, November 2008, 231–252

POST ENUMERATION SURVEY OF THE 2001 PORTUGUESE POPULATION AND HOUSING CENSUSES

Authors: Pedro Simões Coelho
– ISEGI, Universidade Nova de Lisboa, Portugal
psc@isegi.unl.pt

Fernando Casimiro
– Instituto Nacional de Estatística, I.P., Portugal
fernando.casimiro@ine.pt

Received: December 2007 Revised: March 2008 Accepted: April 2008

Abstract:

- Within the framework of the quality control and evaluation program for the Portuguese 2001 Census, the Portuguese statistical office (INE) conducted a Post Enumeration Survey (PES) to measure quality. The main aims of the PES were to evaluate coverage errors and content errors for the target populations. The PES is a probabilistic sampling survey representative at NUTS II level. This paper describes the methodology for this survey. The paper includes a discussion of sample size and allocation resulting from the imposition of maximum coefficients of variation for a set of variables both at regional and national level. The methodology used to obtain predictions for resident populations and dwellings is also presented. These predictions are used in the definition of inclusion probabilities for the primary sampling units. The sampling design is finally compared with two alternative designs (with a smaller number of stages), concluding for the advantage of the proposed design in regard to the survey goals.

Key-Words:

- *Post Enumeration Survey; sampling design; census; population and housing census; census quality; census errors; coverage errors.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

1.1. Context

Quality is an increasingly important subject in the production of statistics. Customers tend to be increasingly demanding and critical about statistical data.

For population and housing censuses, quality evaluation is carried out in various ways, one of them being the Post Enumeration Survey (PES). Usually, when they exist, PES results are assumed to be the final quality indicators for these censuses. "The PES, a special kind of survey designed to measure census coverage and/or content error, has been used effectively in a wide range of countries in recent decades; (...) The final publication should include an estimate of coverage error, together with a full indication of the methods used for evaluating the completeness of the data..." ([20]).

Coverage and content evaluation of population census data are carried out with two main purposes:

(1) to provide customers with quality indicators;

(2) for internal use, to improve knowledge of the problems encountered, in order to improve the capacity to plan and conduct this type of statistical operation in the future.

1.2. Historical note

In the USA, evaluation of population census coverage and content began in 1950, while in Australia "the first PES was run in 1966, but the 1976 PES was the first to be used for population estimates" ([1]). In France, the first quality census evaluation with a PES was carried out in 1962 but the next assessment was not held until 1990 ([6]). Canada started measuring gross undercoverage in 1971, but estimates of net undercoverage are only available from 1991, as 1991 marked the first comprehensive measure of overcoverage following an experimental study in 1986 ([19]). The UK first used a PES to measure census quality in 1961, since when a PES has been conducted in every census ([14, 15, 16, 12]). In other countries the introduction of a PES began later, as in the case of New Zealand where the first one was conducted in 1996.

In fact, post enumeration surveys have been regularly used in countries where a population census is performed, but for 2000 and 2001 some countries

made an even heavier investment in the measurement of census data quality. This is the case of Canada, United Kingdom and the USA ([19, 12, 3, 4]).

The first attempt to conduct a PES in Portugal occurred with the 1981 census, though technical constraints and the lack of human resources did not allow the task to be completed. However, for internal purposes only, a comparative tabulation was made using the two equivalent samples of statistical units (census and PES), which were used to produce two independent but equivalent sets of data tables. For the 1991 census a new PES was designed with strict rules on each person in each selected household being re-interviewed: that is, each person had to be re-interviewed face to face and no other person in the household could be substituted. In Census 91 it was possible to produce gross and net indicators on coverage for each statistical unit but not on the content of census variables. The delay in the census fieldwork and consequently in the PES also made it very difficult to match and apply the automatic rules for the imputation of responses in the PES questionnaires in the same way as for the census responses.

Certain leaders of public opinion also expressed their doubts about the quality of coverage in the 1991 census (about 1% net undercoverage on population, measured by the PES), which led the Portuguese Statistical Office (INE) to decide that the estimates of quality for the 2001 Census should be clear and proven. In order to reach this target, a "special" programme on quality evaluation was designed in which PES was to provide the final quality measure for the 2001 Census.

Given the unexpected outcome from the 1991 Census (the count being approximately 5% below the population estimates made by National Statistical Institute (INE) itself before the first results of the 1991 Census became available), INE was convinced that the 2001 Census would be subject to very close scrutiny by its main users. In a way, despite the fact that ten years have passed since 1991, the 2001 census data would end up being an important evaluation factor of the 1991 census, given that no significant or unexpected demographic "accident" has taken place or was expected to take place in the country's demographic development ([5]).

1.3. The 2001 Portuguese post enumeration survey

The main goal of the census 2001 PES was to evaluate coverage and content errors, giving information to census users about the accuracy of the results, thus allowing to assess the risks involved in basing conclusions or decisions on census data. Coverage and content errors are evaluated for the following universes: buildings, dwellings, private households and resident population.

The evaluation of coverage errors includes three main causes:

(1) Statistical units of the target populations that have not been enumerated;

(2) Statistical units outside the target populations that have been wrongly enumerated;

(3) Statistical units that have been enumerated more than once.

The evaluation of content errors includes census flaws related to observing statistical unit characteristics that can affect the quality of census information about resident population and housing.

To assess coverage and content errors the census enumeration process was repeated in the selected sampling units. At statistical section level recounts of buildings, dwellings, private households and resident population were made. Also, the various types of questionnaires are again completed, for the different statistical units, regarding the characteristics that those units had on census day.

It should be noted that the Portuguese PES presents a number of specificities when compared to other post enumeration surveys, namely: not only aimed to measure coverage errors but also content errors; all measures were obtained using only one sample (though data were obtained from different sampling stages); it was designed within a framework where no sampling frames besides administrative division of the country and auxiliary information regarding population and dwelling estimates were available; it used a three-stage design with selection probabilities proportional to size in the first two sampling stages; it was designed to avoid the selection of sampling units being dependent on the conclusion of census fieldwork in order to reduce the time between the census date and the implementation of the post enumeration survey; it is meant to use information on the geographical coordinates of sampling units in the sampling design; sample size and allocation are obtained by means of an optimization problem that tries to minimize the overall survey cost.

This paper discusses the methodology for this survey. The first section introduces the problem and its context. The sampling design is presented in the second section. Also, the methodology used for defining sample size and allocation between strata, resulting from the imposition of maximum coefficients of variation (CVs) both at regional and national level, for a set of variables, is presented. The following section presents the methodology used in producing predictions for the resident population and dwellings at the time of the census. These predictions are used as auxiliary information in the definition of inclusion probabilities for the primary sampling units. The paper finishes with some final remarks and a discussion about the sampling design (compared with two alternative designs with a smaller number of stages).

2. SAMPLE DESIGN

2.1. Introduction

The quality survey for the 2001 census is a probabilistic sampling survey. It covers the whole of the national territory and aims to be representative at NUTS II[1] level for the variables *dwellings, private households, resident population, active population, employed population, resident population aged 18 years or more and population by decennial age group between 20 and 80 years of age.* Figure 1 shows the partition of Portugal into the 7 NUTS II.

Figure 1: NUTS II division.

A sample of statistical sections[2] is used to evaluate coverage errors for buildings and dwellings, while a sample of dwellings is used to assess coverage errors for private households and resident population and content errors.

[1]NUTS (Nomenclature of Territorial Units for Statistical Purposes) II is an administrative division that divides the country into seven regions (*Norte*, *Centro*, *Lisboa e Vale do Tejo*, *Alentejo*, *Algarve*, *Região Autónoma dos Açores* and *Região Autónoma da Madeira*).

[2]The statistical section is a statistical division corresponding to an area belonging to a single parish (*freguesia*) with approximately 300 dwellings.

The sample is previously stratified by NUTS II. In each stratum a sample of *freguesias*[3], statistical sections and dwellings is obtained. The approach includes the selection, in each stratum, of a multi-stage self-weighted sample through systematic selection with probability proportional to size (pps) at the first and second stage. The primary sampling units are *freguesias*, the secondary sampling units statistical sections and the tertiary sampling units dwellings. The sampling design assures equal probability of selection for dwellings within strata. See [17, pp. 144–150] for general theory about multi-stage designs.

At the first sampling stage inclusion probabilities are defined through the use of auxiliary information based on resident population and estimates of dwelling totals, at census time. An exception is made in the stratum of the *Algarve* where the selection of the sub-sample of *freguesias* is based on resident population estimates in each *freguesia*.

Auxiliary information resulting from preliminary counts from the questionnaire delivery phase of the census is used to define inclusion probabilities for the statistical sections (secondary sampling units). This is due to the impossibility of producing reliable estimates (for population or dwellings) at statistical section level. With this approach it is possible to incorporate updated and high quality auxiliary information in the selection process for statistical sections, which contributes to a more efficient sampling design.

It should be noted that primary unit selection is carried out a priori, i.e. before the census date, using estimates for the number of dwellings and the resident population by *freguesia*. On the other hand, statistical sections are selected as soon as the counts from the questionnaire delivery phase of the census are obtained for each of the previously selected *freguesias*. Given the multistage nature of the sampling design, the selection of statistical sections is not dependent on the conclusion of all counts in the questionnaire delivery phase, but only those referring to *freguesias* selected at the first stage. Such dependence would be undesirable, given the obvious interest in reducing the time between the census date and the quality survey.

Finally, at the third stage, the dwelling samples are extracted, through systematic selection and with equal probabilities as soon as the dwelling recounts are completed for statistical sections selected at the second sampling stage.

A more detailed description is given in the following sections.

[3] *Freguesia* (NUTS V) is an administrative division corresponding to one or more Statistical Sections. At the census day there were 4,208 freguesias in Portugal.

2.2. Selection of *freguesias* (primary units)

At the first sampling stage, *freguesias* are selected in each region (stratum) with probability proportional to the estimated number of dwellings. For the *Algarve* the selection probability is proportional to the estimated resident population. The use of a different approach is motivated by the weak correlation (observed in the simulations using data from the 1991 census) between resident population and dwellings in that region. Therefore, the choice of an alternative sampling design for this region contributes to a significant reduction in sampling effort (cf. Section 2.5).

Freguesias are sorted beforehand using the geographical coordinates of their centroids[4]. In each stratum *freguesias* are ordered by ascending order of their Euclidean distance from the origin. The goal is to assure that the sample is geographically dispersed while still allowing a probability of selection proportional to its size. Finally, *freguesias* are selected through systematic sampling.

The selection probability for freguesia i of stratum h was defined as

$$\pi_{hi} = \begin{cases} \dfrac{A_{hi}}{I_h} & \text{if } A_{hi} < I_h \,, \\ 1 & \text{otherwise} \,, \end{cases} \tag{2.1}$$

where A_{hi} is the estimated number of dwellings (population for the *Algarve*) of *freguesia* i of stratum h.

The selection interval for *freguesias* at stratum h, I_h, is

$$I_h = \frac{A_h}{m_h} \tag{2.2}$$

where m_h is the number of statistical sections to be selected for the sample in stratum h and A_h is the estimated number of dwellings (population for the *Algarve*) in stratum h.

Note that the selection interval I_h is inversely proportional to the number of secondary sampling units, m_h. As it will be explained in more detail in the next section, the reasoning behind this choice is to support the selection of only one secondary sampling unit (*statistical section*) at each primary sampling unit (*freguesia*) with selection probability lower than one.

[4]The point of origin of these coordinates is situated in the Atlantic Ocean to the southwest of Portugal.

2.3. Selection of statistical sections (secondary units)

Lists of statistical sections are formed in *freguesias* selected at the first sampling stage. These sections are sorted using their geographical coordinates (distance from the centroid to the origin).

At the second sampling stage statistical sections are selected through systematic sampling with probability proportional to the number of dwellings obtained in the preliminary counts from the questionnaire delivery phase of the census.

The selection probability for section j of *freguesia* i of stratum h conditioned to the selection of *freguesia* hi is defined as

$$\pi_{hij|hi} = \begin{cases} \dfrac{N_{hij}}{N_{hi}} & \text{if } A_{hi} < I_h \,, \\ \dfrac{A_{hi}}{I_h}\,\dfrac{N_{hij}}{N_{hi}} & \text{otherwise} \,, \end{cases} \tag{2.3}$$

where N_{hij} is the number of dwellings in section j of *freguesia* i of stratum h (data from the preliminary counts from the questionnaire delivery phase of the census), N_{hi} is the number of dwellings at *freguesia* i of stratum h (data from the preliminary counts from the questionnaire delivery phase of the census).

The unconditional selection probability for section j of *freguesia* i of stratum h is consequently

$$\pi_{hij} = \frac{A_{hi}}{I_h}\,\frac{N_{hij}}{N_{hi}} \,. \tag{2.4}$$

To guarantee this selection probability, the selection interval in *freguesia* i of stratum h is defined as

$$\mathbb{I}_{hi} = \begin{cases} N_{hi} & \text{if } A_{hi} < I_h \,, \\ \dfrac{N_{hi}}{A_{hi}}\, I_h = \dfrac{N_{hi}\, A_h}{A_{hi}\, m_h} & \text{otherwise} \,. \end{cases} \tag{2.5}$$

The number of statistical sections being re-enumerated in each *freguesia* selected at the first sampling stage is equal to one for *freguesias* with selection probability lower than one, allowing therefore a strong dispersion of sampled sections in a high number of *freguesias*. The sample size at the second stage will only be higher than one for *freguesias* with selection probability equal to one in order to assure that the unconditional selection probability at the second stage remains proportional to $A_{hi}\, N_{hij}/N_{hi}$.

The number of statistical sections being re-enumerated in *freguesia* i will then be

$$m_{hi} = \begin{cases} 1 & \text{if } A_{hi} < I_h \,, \\ m_h\,\dfrac{A_{hi}}{A_h} & \text{otherwise} \,. \end{cases} \tag{2.6}$$

2.4. Selection of dwellings (tertiary units)

The estimation of coverage errors relative to buildings and dwellings is achieved through the sample of secondary units. For that purpose, each statistical section in the sample should be exhaustively re-enumerated in order to obtain the "true" totals for buildings and dwellings. After obtaining these recounts, a list of dwellings is formed in each statistical section. These lists are used to select the samples of dwellings (tertiary units) to be re-enumerated.

At this third sampling stage dwellings are selected through systematic sampling, with equal probabilities, in order to obtain a self-weighted sample in each stratum.

The selection probability for dwelling k, of section hij, conditioned to the selection of the section to which it belongs, is defined as

$$\pi_{hijk|hij} = \frac{n_{hij}}{N'_{hij}} \tag{2.7}$$

where n_{hij} is the number of sampled dwellings in section hij and N'_{hij} is the number of dwellings in section hij, (obtained from the second stage of the PES).

The unconditional selection probability for dwelling k of section hij is therefore belongs, is defined as

$$\pi_{hijk} = \pi_{hij} \cdot \pi_{hijk|hij} = \frac{A_{hi}}{I_h N_{hi}} \frac{N_{hij}}{N'_{hij}} n_{hij} \ . \tag{2.8}$$

In each section, the sample size for tertiary units is obtained in order to get a self-weighted sample of dwellings in each stratum. For that a constant selection probability is defined in each stratum, equal to the overall sampling rate $f_h = n_h/N'_h$.

The sample size at section j of *freguesia* i in stratum h is consequently given by

$$n_{hij} = \frac{f_h A_h N_{hi} N'_{hij}}{m_h A_{hi} N_{hij}} \ . \tag{2.9}$$

To guarantee the defined selection probabilities, the sampling interval in section hij is defined as

$$I_{hij} = \frac{m_h A_{hi} N_{hij}}{m_h A_h N_{hi}} \ . \tag{2.10}$$

Therefore, the resulting selection probability

$$\pi_{hijk} = \frac{A_{hi}}{I_h N_{hi}} \frac{N_{hij}}{N'_{hij}} n_{hij} = \frac{A_{hi}}{I_h N_{hi}} \cdot \frac{N_{hij}}{N'_{hij}} \cdot \frac{f_h A_h N_{hi} N'_{hij}}{m_h A_{hi} N_{hij}} = f_h$$

will be constant in each stratum.

2.5. Sample size and allocation

It should be remembered that the sample is previously stratified by NUTS II, resulting in seven strata. Also the survey is intended to be representative not only at national level but also at NUTS II level. The overall sample size and its allocation by each stratum was obtained as the solution to the following optimization problem[5]:

$$
\begin{aligned}
(2.11) \qquad & \min \left(C = \sum_{i=1}^{H} c_{1h} m_h + c_{2h} n_h \right) \\
& \text{s.t.} \\
& CV(\hat{\tau}_{k,h}) \leq d_{k,h} \qquad (k = 1, ..., K; \quad h = 1, ..., H) \\
& CV(\hat{\tau}_k) \leq d_k \qquad (k = 1, ..., K) \\
& m_h \leq M_h, \quad n_h \leq N'_h \qquad (h = 1, ..., H) \\
& m_h \geq 0, \quad n_h \geq 0 \qquad (h = 1, ..., H)
\end{aligned}
$$

where K is the number of variables considered for sample dimensioning, H the number of strata, m_h is the sample size of statistical sections in stratum h, n_h is the sample size of dwellings in stratum h, M_h is the number of sections in stratum h, c_{1h} the cost of observing one section in stratum h, c_{2h} the cost of observing one dwelling in stratum h, $CV(\hat{\tau}_{k,h}) = \dfrac{\sqrt{V(\hat{\tau}_{k,h})}}{\tau_{k,h}}$, $CV(\hat{\tau}_k) = \dfrac{\sqrt{V(\hat{\tau}_k)}}{\tau_k}$, $\tau_{k,h}$ is the population total of variable k in stratum h and $\hat{\tau}_{k,h}$ the Horvitz–Thomson estimator for the same parameter. Also, $\tau_k = \sum_{i=1}^{H} \tau_{k,h}$ is the population total of variable k, $\hat{\tau}_k$ is its estimator and $V(\hat{\tau}_k) = \sum_h V(\hat{\tau}_{k,h})$. As proved in Appendix 1, the variance $V(\hat{\tau}_{k,h})$ can be approximated by the expression

$$
\begin{aligned}
V(\hat{\tau}_{k,h}) \;\approx\; & \frac{1}{m_h} \left[\sum_{i \in U_{1h}^I} \frac{A_h A_{hi}}{A_{1h}^2} \left(\frac{A_{1h} \tau_{k,hi}}{A_{hi}} - \tau_{k,1h} \right)^2 \right. \\
& \left. + \sum_{i \in U_{1h}^I} \sum_{j \in U_{hi}^{II}} \frac{A_h N_{hij}}{A_{hi} N_{hi}} \left(\frac{N_{hi} \tau_{k,hij}}{N_{hij}} - \tau_{k,hi} \right)^2 \right] + \frac{N_h'^2}{n_h} \sigma_{k,h,intra}^2 \;,
\end{aligned}
$$

where $\sum_{i \in U_h^I}$ is the summation over all *freguesias* of stratum h, $\sum_{i \in U_{1h}^I}$ is the summation over the *freguesias* of stratum h where $\tau_{a,hi} < I_h$, $\sum_{j \in U_{hi}^{II}}$ is the summation over all the sections of *freguesia* i of stratum h, $\tau_{a,1h}$ is the estimate of the total of dwellings (residents in the *Algarve*) in population U_{1h}^I, $\tau_{k,1h}$ is the total of variable k in the same population and $\sigma_{k,h,intra}^2 = \sum_{i \in U_h^I} \sum_{j \in U_{hi}^{II}} \dfrac{N'_{hij}}{N'_h} \sigma_{k,hij}^2$ is the intra-section variance for variable k in stratum h. All other parameters are as defined in the previous sections.

[5]The problem was solved through Generalized Reduced Gradient Nonlinear Optimization.

The aim of the strategy adopted is to minimize the total sampling cost, C, with the application of maximum limits for the variation coefficients in estimating totals for the K selected variables, at regional level (the stratum corresponding to NUT II) and national level.

The maximum variation coefficients at NUT II level (d_{kh}) were established at 5% for the variables *dwellings, private households, resident population, active population, employed population, and resident population of 18 years of age or more* and at 7% for *resident population by decennial age group between 20 and 80 years of age*. Also, the maximum variation coefficients for estimating national totals (d_k) were set at 3% for the variables *dwellings, private households, resident population, active population, employed population, and resident population of 18 years of age or more* and at 3.5% for *resident population by decennial age group between 20 and 80 years of age*.

Figure 2 shows the geographical location of sections in the sample. Attention should be paid to the location of sample sections associated with high population concentrations in coastal areas and urban centers. The calculations used to determine sample sizes were based on data from the 1991 census[6].

It should be remembered that a specific sampling design was adopted in the Algarve, since the selection probabilities for primary sampling units in that NUTS II were defined as proportional to the estimated resident population. If the sampling design in the *Algarve* were the same as that adopted in other NUTS II, using the number of dwellings to define selection probabilities, the necessary sample size to guarantee the achieved variation coefficients would be equal to 122 statistical sections. This result clearly demonstrates the advantage of the procedure adopted, which is justified by the low correlation between resident population and dwellings in that region (cf. Table 1). In fact, one should remember that the *Algarve* is a tourist region where many people keep a second home.

It should also be noted that the high sample size obtained in *Lisboa e Vale do Tejo* is essentially justified by the significant variance that some of the variables show at *freguesia* level, as well as by the low correlation between resident population and dwellings at section level. This sample could be downsized using a design similar to that adopted in the *Algarve*[7]. The decision to keep the primary

[6]Since the simulation used 1991 census data summarized at *freguesia* level, the variances ($k = 1, ..., K;\ h = 1, ..., H$) were replaced by estimates obtained from other surveys conducted by INE. In practice, c_{1h} and c_{2h} were considered non invariant with h, i.e. $c_{1h} = c_1$; $c_{2h} = c_2$, $h = 1, ..., H$. A new set of restrictions was also imposed, $f_h = f$, $h = 1, ..., H$, in order to achieve a constant selection probability for all dwellings in the country. The overall sampling rate f came out approximately equal to 0.00275. Simulations showed that the impact of this procedure on the overall sampling cost was moderate.

[7]In the simulations using data from the 1991 census the reduction in sample size (for the same level of precision) would be from 110 to 97 statistical sections.

units proportional to the number of dwellings in that region resulted from the observation that the reduction in sample size would be less significant than in the *Algarve*. In addition, the strategy adopted shows certain advantages in field procedures, associated with the greater stability of the sample size at the last sampling stage.

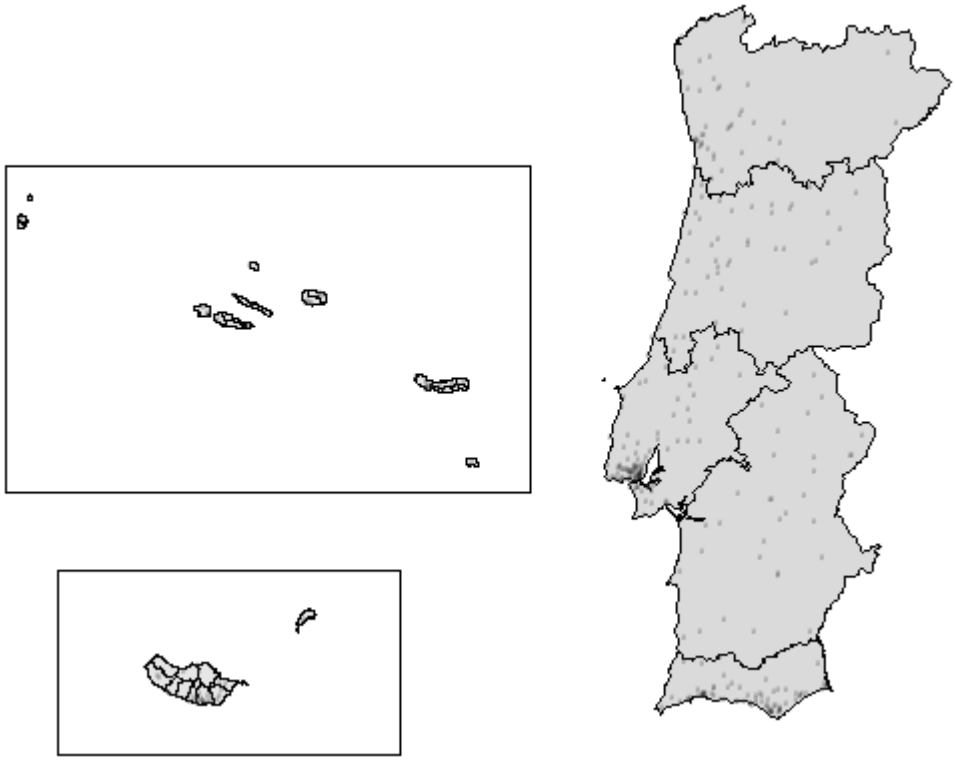

Figure 2: Statistical section in the sample.

Table 1: Standard deviations and correlations between dwellings and resident population (data from the 1991 census).

Stratum (NUTS II)	Freguesia			Statistical Section		
	Correlation	Standard deviation (dwellings)	Standard deviation (population)	Correlation	Standard deviation (dwellings)	Standard deviation (population)
Norte	0.99	1,286	3,532	0.84	135	422
Centro	0.97	992	2,283	0.75	99	278
LVT	0.98	3,936	9,608	0.57	84	249
Alentejo	0.99	795	1,849	0.75	82	212
Algarve	0.87	3,432	5,273	0.43	115	279
Açores	0.96	435	1,477	0.68	82	320
Madeira	0.98	1,350	5,131	0.61	73	318

3. DWELLING AND RESIDENT POPULATION ESTIMATES

It has already been mentioned that the selection probabilities at the first sampling stage use estimates of the number of dwellings and the resident population by *freguesia*, in relation to the census date.

The production of these estimates is addressed below, particularly as regards the data sources and methodology used. The proposed estimators were tested with the exhaustive observation of 107 *freguesias* in April 2000. This observation followed a similar approach to the one adopted in the 2001 census. Estimators for *freguesia* totals (obtained using the methodology presented in the following sections) showed a mean absolute relative error of 11% in estimating the resident population of *freguesias* in the test and of 8% in estimating the number of dwellings. Note that these errors are compatible with the approach used for determining sample size, since that approach was based on data from the 1991 census where the mean absolute relative error for the estimated number of dwellings was about 12%. A more detailed presentation of the methodology used can be found in [7].

3.1. Estimation of the number of dwellings by *freguesia*

Some of the data sources available for the number of dwellings were: the results of the 1991 census, statistics from the INE construction survey and EDP household registers (Portuguese electricity company data on domestic consumption locations).

The estimate for the number of dwellings in freguesia i on the census date is obtained as

$$A_{hi} = 0.5 \times C^{hi} + 0.5 \times D^{hi}_{2001} \ , \tag{3.1}$$

where, A_{hi} is the estimate of the number of dwellings in *freguesia* hi, at the census date, C^{hi} is the simple count of the number of domestic electricity consumption contracts in the EDP registers, by *freguesia*, D^{hi}_{2001} is an estimate derived from accumulating the annual balance of newly constructed dwellings and demolished dwellings (data from the construction survey), on the basis of the number of dwellings obtained in the 1991 census (corrected from the coverage error).

This last estimate is given by[8]:

$$\begin{aligned} D^{hi}_t &= D^{hi}_{t-1} + NewD^{hi}_t - Dem^{hi}_t \\ D^{hi}_{91} &= D^{hi}_{C91} \end{aligned} \tag{3.2}$$

[8]This methodology presented various limitations, namely: the absence of information about demolitions prior to 1993; the fact that data prior to 1994 is only available at *concelho* level

where D_t^{hi} is the estimate of the number of dwellings in *freguesia hi* at year t, $NewD_t^{hi}$ is the number of dwellings constructed in *freguesia hi*, at year t (data from the construction survey), Dem_t^{hi} is the number of demolitions in *freguesia hi*, at year t (data from the construction survey), D_{C91}^{hi} is the number of dwellings in *freguesia hi* counted in the 1991 census.

The decision to use an average of the two estimates is based on the observation (using the 107 test *freguesias*) that EDP data tends to systematically underestimate the true number of dwellings, while the estimate obtained from the construction survey shows a tendency of to overestimate them. In the absence of other information on the precision of each of the alternative estimators, a natural choice in producing a weighted average of both estimates is to use equal weighting.

3.2. Estimation of resident population by *freguesia*

Some of the data sources available on resident population were: the results of the 1991 census, official data on births and deaths in the decade, electoral roll databases, and the legalization of immigrants and residence cancellations (data from the Immigration Service).

The estimates produced result from the accumulation of the balance between births and deaths, the balance of transfers in the electoral rolls, and net immigration (taken as the difference between legalization requests and cancellations in the Immigration Service). The resident population enumerated in the 1991 census (corrected from the coverage error) was taken as a base for the calculation.

The estimate of the resident population in *freguesia hi*, at year t, is

$$\begin{aligned} Pop_t^{hi} &= Pop_{t-1}^{hi} + BB_t^{hi} + TrB_t^{hi} + FmB_t^{hi} \\ Pop_{91}^{hi} &= Pop_{C91}^{hi} \end{aligned} \tag{3.3}$$

where

- $BB_t^{hi} = Births_t^{hi} - Deaths_t^{hi}$;
- $Births_t^{hi}$ is the number of births in *freguesia hi*, at year t;
- $Deaths_t^{hi}$ is the number of deaths in *freguesia hi*, at year t;

(aggregation of *freguesias*); and the unavailability of data posterior to 1999. These limitations were overcomed by carrying out, respectively, an estimation of demolitions for 1991 and 1992 based on information about the following years, allocation of data prior to 1994 by *freguesia* using the average structure of each *concelho* in the period 1994–1999 and the prediction of the series value at the census day by adjustment of a linear regression model.

- $TrB_t^{hi} = T_t^{hi} - E_t^{hi}$;
  - T_t^{hi} is the number of transfers to *freguesia hi*, at year t, in the electoral roll;
  - E_t^{hi} is the number of cancellations due to transfer from *freguesia hi*, at year t, in the electoral roll;

- $FmB_t^{hc} = LR_t^{hc} - C_t^{hc}$, $FmB_t^{hi} = FmB_t^{hc} \frac{FPop_{C91}^{hi}}{FPop_{C91}^{hc}}$;
  - FmB_t^{hc} is the net migration of foreigners to *concelho* (municipality) *hc*, at year t;
  - LR_t^{hc} is the number of legalization requests from foreigners in *concelho hc*, at year t;
  - C_t^{hc} is the number of residence permit cancellations for foreigners in *concelho hc*, at year t;
  - FmB_t^{hi} is the net migration of foreigners to *freguesia hi*, at year t;
  - $FPop_{C91}^{hi}$ is the resident population of foreigners in *freguesia hi*, recorded in the 1991 census;
  - $FPop_{C91}^{hc}$ is the resident population of foreigners in *concelho hc*, recorded in 1991 census.

The estimated resident population for the census date is then

$$P_{hi} = Pop_{2001}^{hi} \ . \tag{3.4}$$

Thus, the estimator includes information about births and deaths, internal migrations and foreign immigration. It may be presumed that internal migration and foreign immigration had a great impact on the resident population at *freguesia* level, since there were less than 90,000 persons as natural increase at national level between 1991 and 2001.

It should also be noted that, although data for births and deaths are considered to be totally reliable, this is not the case with migration, both internal and international. For this reason, transfers in the electoral rolls were taken as a proxy for internal migration at *freguesia* level. In fact, the impossibility of producing reliable estimates for internal migrations at this aggregation level motivated the search for a variable that could be considered a proxy for internal migration as it was reliable at *freguesia* level. The main limitation is the fact that migrations of people under 18 years of age are not included. In addition, calculation of the estimates demanded the allocation of the net foreign migration (only available at *concelho* level) to *freguesias*, using the structure observed in the 1991 census.

4. DISCUSSION

The PES was designed to evaluate coverage errors and content errors in the main statistical units: buildings, dwellings, private households and resident population. For this purpose the enumeration process was repeated in the selected sampling units. The various questionnaires were completed again, for the different statistical units, with the characteristics that those units had at the time of the census day.

This paper discussed the methodology used for this survey, which is based on a three-stage sample. Primary and secondary sampling units were selected with probability proportional to the size. For this purpose, at the first sampling stage inclusion probabilities were defined through the use of auxiliary information based on the estimated resident population and dwelling totals, at the time of the census. Auxiliary information, resulting from preliminary counts obtained at the questionnaire delivery phase in the census, was used in the definition of inclusion probabilities for secondary sampling units. With this approach it was possible to incorporate updated and high-quality auxiliary information in the selection process for the statistical sections, contributing to a more efficient sampling design. In the first two sampling stages auxiliary information regarding the geographical coordinates of area units (statistical section and *freguesia*) was used in order to obtain an implicit stratification. Particularly with regard to the number of stages, the design took operational restrictions into account. The goal was to make use of available auxiliary information to determine the appropriate selection probabilities and to avoid the sample selection being totally dependent on the conclusion of all counts in the questionnaire delivery phase. In fact, the lowest level of aggregation for which information used to define inclusion probabilities is available is the *freguesia* level. So, the selection of statistical sections as primary sampling units would make it impossible to use reliable auxiliary information in defining such probabilities.

Sample size and allocation between strata were obtained as the solution to an optimization problem that minimizes the total survey cost, with maximum limits for the coefficients of variation in estimating totals for a number of variables, both at regional level (the stratum corresponding to NUTS II) and at national level. This approach resulted in a sample of 367 statistical sections.

Nevertheless, alternative sampling designs, with a smaller number of stages, could have been conceived, in order to try to reduce a possible design effect. This alternative approach could then be based on a two-stage sampling, successively selecting statistical sections and dwellings.

One possible approach would then be to select statistical sections with equal probability. A two-stage sampling with selection of statistical sections with equal probability at the first stage was simulated using data from the 1991 census.

To guarantee the same precision, that design would have to be based on the sample sizes shown in the last column of Table 2 [9]. It should be noted that with the exception of *Lisboa e Vale do Tejo*, the sample sizes necessary to achieve the same variation coefficients would be substantially higher than the ones used in the proposed design, resulting in an impressive increase in the overall sample size from 267 sections to 743 sections.

Table 2: Sample sizes for three alternative sampling desig.

Stratum (NUTS II)	**Sample Size**		
	3 stages	**2 stages Selection of sections (with probability proportional to the number of dwellings)**	**2 stages Selection of sections (with equal probability)**
Norte	42	52	159
Centro	46	63	119
Lisboa e Vale do Tejo	110	87	97
Alentejo	33	55	93
Algarve	73	122	145
Açores	32	40	74
Madeira	31	38	56
Total	367	457	743

Another approach would be to select statistical sections, at the first sampling stage, with probability proportional to preliminary counts obtained from the questionnaire delivery phase in the 2001 census. With such an approach the selection of sections would be dependent on the conclusion of all the counts from the questionnaire delivery. This dependency is undesirable given the goal of reducing the period between the census date and the implementation of the post enumeration survey. Moreover, certain operational restrictions recommend that the geographical distribution of the sample should be known before the selection of the sample is possible.

Furthermore, this alternative design would not lead to a more precise estimation. Results from the simulation with data from the 1991 census (cf. Table 2), corresponding to a two stage sampling with selection of statistical sections with probability proportional to the preliminary counts from the questionnaire delivery at the first stage, show that the design adopted is generally more efficient than this alternative design[10]. In fact, it can be seen that the sample size necessary

[9]For this propose, in each stratum, the total estimator variance was approximated by the expression $V(\hat{\tau}_{k,h}) \approx \frac{M_h^2}{m_h(m_h-1)} \sum_{i \in U_h^I} \sum_{j \in U_{hi}^{II}} \left(\tau_{k,hij} - \frac{\tau_{k,h}}{M_h}\right)^2$, where M_h represents the number of sections in stratum h. This implies that for this benchmark, sampling error due to the last sampling stage is ignored.

[10]In the simulation the total estimator variance, in each stratum, was approximated by the expression $V(\hat{\tau}_{k,h}) \approx \frac{1}{m_h} \sum_{i \in U_h^I} \sum_{j \in U_{hi}^{II}} \frac{N_{hij}}{N_h} \left(\frac{N_h \tau_{k,hij}}{N_{hij}} - \tau_{k,h}\right)^2$.

to achieve the same variation coefficients would in general be greater than that obtained in the design adopted. It should be noted that this increase in sampling effort would be particularly significant in the *Algarve* because it would no longer be possible to use estimates for resident population in the definition of inclusion probabilities. In order to achieve the same precision this alternative design would cause an increase in the overall sample size from 267 sections to 457 sections[11].

Assuming that the cost of observing one statistical section is not significantly different among the three designs considered, it can be concluded that the survey cost (associated with the first sampling stage) would increase by 25% for the two-stage design with probability proportional to the number of dwellings and about 100% for the two stage design with equal probabilities. Only in *Lisboa e Vale do Tejo* would the two-stage designs lead to a reduction in sample size. This is due to the fact that the variance in *freguesia* totals as regards dwellings and population is higher in this stratum, while the same does not hold at section level. In addition, this was the region where dwelling estimates showed the poorest precision, affecting the quality of selection probabilities at the first stage of the three-stage design.

In fact, in the proposed design, the number of sections selected in each *freguesia* selected for the sample at the first stage is usually equal to one. Only in some *freguesias*, with inclusion probabilities equal to one, will more than one statistical section be selected for the sample in order to keep the unconditional selection probability at the second stage proportional to size. In this way it is possible to avoid a high concentration of sampled sections within a small number of *freguesias* and the typically associated design effect. Moreover it can be observed from Table 1 that the correlation between dwellings and resident population is significantly higher at *freguesia* level than at section level. This means that selection probabilities are more closely correlated with resident population totals at the first sampling stage than at the second, which can be considered as an indication of the superiority of the three-stage design.

Furthermore, a methodology for producing predictions for the resident population and dwellings at the time of the census was presented. This was achieved by combining demographic equations with information from other sources (data from other national surveys, data from the Portuguese electricity company on domestic consumption locations, data from the electoral rolls and data from the Immigration Service on the legalization of immigrants). These predictions were used as auxiliary information for defining inclusion probabilities for the primary sampling units. From a test carried out in 2000, it was concluded that their precision (a mean absolute relative error of 11% in estimating the resident population of *freguesias* in the test and of 8% in estimating the number of dwellings) was compatible with the aims of the survey design.

[11]The simulations were based on the assumption that the dwelling counts obtained from the questionnaire delivery phase of the census are free of error. In the (probable) situation where this assumption does not hold, the alternative two-stage design could erode precision even further.

APPENDIX

A. DERIVATION FOR THE APPROXIMATE VARIANCE OF $\hat{\tau}_{k,h}$

It should be noted that is obtained with a three-stage sampling design. Its approximate variance (Särndal, *et al.* 1992, pp. 148–149) can be written as

$$\begin{aligned} V(\hat{\tau}_k, h) \;\approx\; & \sum_{i\in U_h^I} \sum_{i'\in U_h^I} (\pi_{i,i'} - \pi_i \pi_{i'}) \, \frac{\tau_{k,hi}}{\pi_i} \, \frac{\tau_{k,hi'}}{\pi_{i'}} \\ & + \sum_{i\in U_h^I} \sum_{j\in U_{hi}^{II}} \sum_{j\in U_{hi}^{II}} \frac{\pi_{ij,ij'|i} - \pi_{ij|i}\,\pi_{ij'|i}}{\pi_i} \; \frac{\tau_{k,hij}}{\pi_{ij|i}} \; \frac{\tau_{k,hij'}}{\pi_{ij'|i}} \\ & + \sum_{i\in U_h^I} \sum_{j\in U_{hi}^{II}} \frac{V(\hat{\tau}_{k,hij})}{\pi_{ij}} \end{aligned}$$

where U^I and U^{II} represent respectively the population of primary units and secondary units.

Consider in each stratum h, two subpopulations: U_{1h}^I represents primary units of stratum h, such as $A_{hi} < I_h$, and U_{2h}^I the population formed by primary units in stratum h, where $A_{hi} \geq I_h$.

The sampling design is also such that $m_{hi} = 1$, $\forall\, i \in U_{1h}^I$ and $E(m_{hi}) = \frac{A_{hi}}{I_h}$, $\forall\, i \in U_{2h}^I$, with m_{hi} being the size of the sub-sample of sections corresponding to *freguesia* i of stratum h. Using an approximation through a sampling design with replacement we have

$$V(\hat{\tau}_{k,h}) \;\approx$$

$$\begin{aligned} \approx\; & \frac{1}{m_{1h}} \left[\sum_{i\in U_{1h}^I} \frac{A_{hi}}{A_{1h}} \left(\frac{A_{1h}\,\tau_{k,hi}}{A_{hi}} - \tau_{k,1h} \right)^2 + \sum_{i\in U_{1h}^I} \sum_{j\in U_{1hi}^{II}} \frac{A_{1h}\,N_{hij}}{A_{hi}\,N_{hi}} \left(\frac{N_{hi}\,\tau_{k,hij}}{N_{hij}} - \tau_{k,hi} \right)^2 \right] \\ & + \sum_{i\in U_{1h}^I} \sum_{j\in U_{1hi}^{II}} \frac{A_{1h}\,N_{hi}}{m_{1h}\,A_{hi}\,N_{hij}} V(\hat{\tau}_{k,hij}) + \frac{1}{m_h} \sum_{i\in U_{2h}^I} \sum_{j\in U_{2hi}^{II}} \frac{A_h\,N_{hij}}{A_{hi}\,N_{hi}} \left(\frac{N_{hi}\,\tau_{k,hij}}{N_{hij}} - \tau_{k,hi} \right)^2 \\ & + \sum_{i\in U_{2h}^I} \sum_{j\in U_{2hi}^{II}} \frac{A_h\,N_{hi}}{m_h\,A_{hi}\,N_{hij}} V(\hat{\tau}_{k,hij}) \end{aligned}$$

$$\begin{aligned} =\; & \frac{1}{m_h} \left[\sum_{i\in U_{1h}^I} \frac{A_h\,A_{hi}}{A_{1h}^2} \left(\frac{A_{1h}\,\tau_{k,hi}}{A_{hi}} - \tau_{k,1h} \right)^2 + \sum_{i\in U_h^I} \sum_{j\in U_{hi}^{II}} \frac{A_h\,N_{hij}}{A_{hi}\,N_{hi}} \left(\frac{N_{hi}\,\tau_{k,hij}}{N_{hij}} - \tau_{k,hi} \right)^2 \right] \\ & + \frac{N_h'^2}{n_h} \, \sigma_{k,h,intra}^2 \end{aligned}$$

where m_h is the size of the sample of sections corresponding to stratum h, $m_{1h} = m_h \dfrac{A_{1h}}{A_h}$, $V(\hat{\tau}_{k,hij}) = \dfrac{N'_{hij}\, m_h\, A_{hi}\, N_{hij}}{f_h\, A_h\, N_{hi}}\, \sigma^2_{k,hij}$, $\sigma^2_{k,h,intra} = \displaystyle\sum_{i \in U_h^I} \sum_{j \in U_{hi}^{II}} \frac{N'_{hij}}{N'_h}\, \sigma^2_{k,hij}$ is the intra-section variance in stratum h for variable k, and $\sigma^2_{k,hij}$ is the population variance of dwelling totals for variable k in section hij.

REFERENCES

[1] Australian Bureau of Statistics (1995). *Demographic Estimates and Projections: Concepts, Sources and Methods, Chapter 4 (Population Estimates: Data Sources – Census Undercount)*, Australian Bureau of Statistics, Australia.

[2] Australian Bureau of Statistics (1999). *Measuring Census Undercount in Australia and New Zealand (Working Paper No. 99/04)*, Australian Bureau of Statistics, Australia.

[3] Bureau of the Census (2000a). *Census 2000, A.C.E. Methodology*, Vol. 1, Bureau of the Census, USA.

[4] Bureau of the Census (2000b). *The Design of The Census 2000 Accuracy and Coverage Evaluation*, Bureau of the Census, USA.

[5] Casimiro, F. (1998). A Avaliação da Qualidade nos Recenseamentos da População e Habitação em Portugal, *Revista de Estatística*, **2**, 103–107.

[6] Coeffic, N. (1993). L'Enquête Post-Censitaire de 1990, Une mesure de l'exhaustivité du recensement, *Population*, **6**, 1665–1682.

[7] Coelho, P.S. (2001). Modelos de Estimação para os Censos 2001, *Actas do Seminário Censos 2001*, Instituto Nacional de Estatística, Lisbon.

[8] Dixie, J. (2000). *Planning for The 2001 Census Coverage Survey in England and Wales*, Office for National Statistics, UK.

[9] Gabinete dos Censos 2001 (2001). *Censos 2001 Programa Global*, Instituto Nacional de Estatística, Lisbon.

[10] Hansen, M.; Urvitz, W. and Madow, W. (1953). *Sample Survey Methods and Theory*, Vol. I and II, Wiley, New York.

[11] Instituto Nacional de Estatística (1991). *Censos 91 – Inquérito de Qualidade / Manual de Instruções*, Instituto Nacional de Estatística, Lisbon.

[12] Office for National Statistics (2000). *A One Number Census, Census Consultation Paper*, Government Statistical Service, UK.

[13] Office for National Statistics (2001). *Census 2001, A Guide to The One Number Census*, Government Statistical Service, UK.

[14] Office of Population Censuses and Surveys (1968). *Census 1961 General Report*, Office of Population Censuses and Surveys, Great-Britain.

[15] Office of Population Censuses and Surveys (1983). *Census 1971 General Report*, Office of Population Censuses and Surveys, Great-Britain.

[16] Office of Population Censuses and Surveys (1990). *Census 1981 General Report*, Office of Population Censuses and Surveys, Great-Britain.

[17] Särndal, C.E.; Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

[18] Sharp, J.; Kindelberger, J.C. and Bushery, J. (2000). *Measuring Response Variance in Census 2000*, Bureau of the Census, USA.

[19] Statistics Canada (1999). *1996 Census Technical Reports – Coverage (Catalogue No. 92-370-XIE)*, Ministry of Industry, Ottawa.

[20] United Nations (1999). *Principles and Recommendations for Population and Housing Censuses*, Revision 1, United Nations.

REVSTAT – Statistical Journal
Volume 6, Number 3, November 2008, 253–277

ESTIMATION AND FORECASTING IN SUINAR(1) MODEL

Authors: Nélia Silva
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
neliasilva@ua.pt

Isabel Pereira
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
isabel.pereira@ua.pt

M. Eduarda Silva
– Faculdade de Economia da Universidade do Porto, UI&D Matemática e Aplicações da Universidade de Aveiro, Portugal
mesilva@fep.up.pt

Received: April 2008 Revised: September 2008 Accepted: October 2008

Abstract:

- This work considers a generalization of the INAR(1) model to the panel data first order Seemingly Unrelated INteger AutoRegressive Poisson model, SUINAR(1). It presents Bayesian and classical methodologies to estimate the parameters of Poisson SUINAR(1) model and to forecast future observations of the process. In particular, prediction intervals for forecasts — classical approach — and HPD prediction intervals — Bayesian approach — are derived. A simulation study is provided to give additional insight into the finite sample behaviour of the parameter estimates and forecasts.

Key-Words:

- *Forecasts; Gibbs sampling; INAR model; panel data.*

AMS Subject Classification:

- 62CF15, 62M10, 62M20.

1. INTRODUCTION

The usual linear models for time series have been used successfully for modelling stationary dependent sequences under the assumption of Gaussianity, which is inappropriate for modelling counting processes. Motivated by the need of modelling correlated series counts, the INteger-valued AutoRegressive (INAR) process was proposed by Al-Osh and Alzaid (1987) and Mckenzie (1985). The INAR model has been extensively studied in the literature and successfully applied in different contexts. A generalization of the INAR model to the multivariate case has been considered by Latour (1997). Here, our interest lies in models for integer-valued panel data, which are a particular case of multivariate data. The simplest such model is considered in Silva *et al.* (2005) and consists of independent replicates of the INAR model. However, in many practical situations, namely in econometric data, the individuals are not uncorrelated. Such an example is the panel data of entry and exit of plants in Swedish municipalities considered by Berlung and Brannas (1996). To model these data, the authors propose a multivariate integer-valued INAR(1) model related to the Seemingly Unrelated Regression model, SUR, as follows.

Consider a panel of integer-valued data consisting of r individuals and $n-1$ time periods, $X_{k,t}$, $k=1,...,r$, $t=2,...,n$, satisfying the following r variate Poisson INAR(1) model with parameters which are constant along the time but different from individual to individual,

$$X_{k,t} = \alpha_k \circ X_{k,t-1} + \epsilon_{k,t} \ , \qquad k=1,...,r, \quad t=2,...,n \ , \tag{1.1}$$

where $x_{k,1}$ is known, $\alpha_k \circ X_{k,t-1} | X_{k,t-1} \sim B(X_{k,t-1}, \alpha_k)$, $\alpha_k \in (0,1)$, $\epsilon_{k,t}$ are, for each $k=1,...,r$, Poisson random variables with parameter μ_k and, moreover, $\epsilon_{k,t}$ and $X_{k,t-1}$ are independent, for all k and t.

The dependence between individuals is modelled in (1.1) through the innovations term by

$$\epsilon_{k,t} = \epsilon^*_{k,t} + \zeta_t \ , \qquad k=1,...,r, \quad t=2,...,n \ .$$

Thus, equation (1.1) takes the form

$$X_{k,t} = \alpha_k \circ X_{k,t-1} + \epsilon^*_{k,t} + \zeta_t \ , \qquad k=1,...,r, \quad t=2,...,n \ , \tag{1.2}$$

with $\epsilon^*_{k,t} \sim P(\lambda_k)$ i.i.d., $k=1,...,r$; $\zeta_t \sim P(\delta)$ i.i.d., $t=2,...,n$; $\epsilon^*_{k,t}$ and ζ_t are independent for $k=1,...,r$, $t=2,...,n$.

The model defined in (1.2) is called *Seemingly Unrelated INteger Auto-Regressive*, SUINAR, since the individuals appear independent from each other.

Particular situations of the model defined in (1.1) were studied by Silva *et al.* (2005) — PoRINAR(1) model — where the parameters are constant along the time and from individual to individual, i.e., considering independent replicates of the PoINAR(1) model. Berglund and Brännäs (2001), Blundell *et al.* (1999) and Böckenholt (1999), considered a generalization of this model in which the parameters depend on exogenous variables and vary with time and from individual to individual.

In time series analysis we are usually interested in estimating the underlying model and in the predictive capabilities of that model. Thus, the aim of this study is to establish a comparison between classical and Bayesian approaches in order to conduct inference for model parameters and obtain predictions for future values. The remaining of the paper is organized as follows. In Section 2, the SUINAR process is introduced and some properties of the model are derived. In Section 3, the estimation of the parameters is studied under several classical methods and Bayesian methodology which requires the use of an MCMC algorithm — ARMS — for which we give full details. In Section 4, forecasts of future observations and prediction intervals are derived, under both approaches. In Section 5, the results are illustrated through a simulation study. Finally, in Section 6 some concluding remarks are given.

2. THE SUINAR(1) MODEL AND ITS PROPERTIES

Equation (1.2) is written in matrix form as

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix}_t = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_r \end{bmatrix} \circ \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_r \end{bmatrix}_{t-1} + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \vdots \\ \epsilon_r^* \end{bmatrix}_t + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_t \zeta_t \ ,$$

or alternatively

$$\mathbf{x}_{\cdot t} \ = \ \mathbf{A} \circ \mathbf{x}_{\cdot (t-1)} + \boldsymbol{\epsilon}_{\cdot t} + \mathbf{1}_r \, \zeta_t \ , \qquad t = 2, ..., n \ ,$$

with

$$\mathbf{A} \circ \mathbf{x}_{\cdot (t-1)} \ = \ \left(\alpha_1 \circ X_1 = \sum_{i=1}^{X_1} B_{i1}, \ \ \ldots, \ \ \alpha_r \circ X_r = \sum_{i=1}^{X_r} B_{ir} \right)'_{t-1} \ ,$$

where $\mathbf{x}_{\cdot t} = (X_{1,t}, X_{2,t}, ..., X_{r,t})$, B_{ik} are i.i.d. Bernoulli random variables with α_k as the success probability and independent of $\mathbf{x}_{t-1}$ and ϵ_t, $t = 2, ..., n$.

The following properties are important for the remainder of the paper.

1. Let $\boldsymbol{\epsilon}_{\cdot t} = \boldsymbol{\epsilon}^*_{\cdot t} + \zeta_t \mathbf{1}_r$. The covariance matrix of $(\boldsymbol{\epsilon}_{\cdot t})$ at lag j is given by

$$\begin{aligned}\boldsymbol{\gamma}_\epsilon(j) &= \mathrm{Cov}\big(\boldsymbol{\epsilon}_{\cdot t}, \boldsymbol{\epsilon}_{\cdot (t+j)}\big) \\ &= \begin{bmatrix} \mathrm{cov}(\epsilon_{1,t}, \epsilon_{1,t+j}) & \mathrm{cov}(\epsilon_{1,t}, \epsilon_{2,t+j}) & \cdots & \mathrm{cov}(\epsilon_{1,t}, \epsilon_{r,t+j}) \\ \mathrm{cov}(\epsilon_{2,t}, \epsilon_{1,t+j}) & \mathrm{cov}(\epsilon_{2,t}, \epsilon_{2,t+j}) & \cdots & \mathrm{cov}(\epsilon_{2,t}, \epsilon_{r,t+j}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\epsilon_{r,t}, \epsilon_{1,t+j}) & \mathrm{cov}(\epsilon_{r,t}, \epsilon_{2,t+j}) & \ldots & \mathrm{cov}(\epsilon_{r,t}, \epsilon_{r,t+j}) \end{bmatrix}.\end{aligned}$$

When $j = 0$, it follows that

$$\boldsymbol{\gamma}_\epsilon(0) = \begin{bmatrix} \lambda_1 + \delta & \delta & \cdots & \delta \\ \delta & \lambda_2 + \delta & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & \lambda_r + \delta \end{bmatrix}.$$

If $j \geq 1$, then $\gamma_\epsilon(j) = 0$, due to the independence between $\epsilon^*_{k,t}$ and ζ_t for $k = 1, ..., r,\ t = 2, ..., n$.

2. The mean value of the process $\mathbf{x}_{\cdot t}$ is given by

$$\mathrm{E}(\mathbf{x}_{\cdot t}) = (\mathbf{I}_r - \mathbf{A})^{-1} (\boldsymbol{\lambda} + \delta \mathbf{1}_r),$$

where $\mathbf{x}_{\cdot t} = (X_{1,t}, X_{2,t}, ..., X_{r,t})$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_r)$ and $\mathbf{I}_r$ is the $(r \times r)$ identity matrix.
For the k-th individual, we have

$$\mathrm{E}[X_{k,t}] = (\lambda_k + \delta)/(1 - \alpha_k), \qquad k = 1, ..., r.$$

3. The covariance matrix of the process, $\mathbf{x}_{\cdot t}$, is defined by

$$\boldsymbol{\gamma}_X(0) = \begin{bmatrix} (\lambda_1+\delta)/(1-\alpha_1) & \delta/(1-\alpha_1\alpha_2) & \cdots & \delta/(1-\alpha_1\alpha_r) \\ \delta/(1-\alpha_2\alpha_1) & (\lambda_2+\delta)/(1-\alpha_2) & \cdots & \delta/(1-\alpha_2\alpha_r) \\ \vdots & \vdots & \ddots & \cdots \\ \delta/(1-\alpha_r\alpha_1) & \delta/(1-\alpha_r\alpha_2) & \cdots & (\lambda_r+\delta)/(1-\alpha_r) \end{bmatrix}. \tag{2.1}$$

4. The covariance matrix $\mathbf{x}_{\cdot t}$ at lag j is given by

$$\boldsymbol{\gamma}_X(j) = \mathrm{E}\Big[\big(\mathbf{x}_{\cdot t} - \mathrm{E}(\mathbf{x}_{\cdot t})\big)\big(\mathbf{x}_{\cdot (t-j)} - \mathrm{E}(\mathbf{x}_{\cdot (t-j)})\big)\Big]' = \mathbf{A}^j \boldsymbol{\gamma}_X(0), \quad j = 1, 2, \ldots .$$

3. PARAMETER ESTIMATION

In this section we consider the estimation of the $2r+1$ unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta) = (\alpha_1, \alpha_2, ..., \alpha_r;\ \lambda_1, \lambda_2, ..., \lambda_r;\ \delta)$ of the SUINAR(1) process from the sample $\mathbf{x}_{r,n} = \big\{X_{k,t};\ k = 1, 2, ..., r;\ t = 1, 2, ..., n\big\}$. The methods under study are the Conditional Maximum Likelihood, Conditional Least Squares, Method of Moments and Bayesian methodology.

3.1. Classical Approach

3.1.1. Conditional Maximum Likelihood Estimators

The likelihood function, conditional on $\mathbf{x}_{\cdot 1} = (x_{1,1}, x_{2,1}, ..., x_{r,1})$, is given by the following expression

$$
\begin{aligned}
L\big(\mathbf{x}_{r,n}; \boldsymbol{\theta} | \mathbf{x}_{\cdot 1}\big) &= \prod_{k=1}^{r} \prod_{t=2}^{n} P\big(X_{k,t} = x_{k,t} \,|\, X_{k,t-1} = x_{k,t-1}\big) \\
&= \prod_{k=1}^{r} \prod_{t=2}^{n} \sum_{i=0}^{M_{k,t}} \exp\big[-(\lambda_k+\delta)\big] \frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i} ,
\end{aligned} \tag{3.1}
$$

with $M_{k,t} = \min(x_{k,t}, x_{k,t-1})$.

Estimates for δ and λ_k, $k=1, ..., r$, cannot be obtained separately due to the term $(\lambda_k+\delta)^{x_{k,t}-i}$. Thus, we consider $\mu_k = \lambda_k + \delta$ in the expression (3.1), and we obtain the conditional maximum likelihood (CML) estimates of α_k and μ_k.

The CML estimates satisfy the following system, where the equations are obtained by cancelling the derivatives of the logarithm of expression (3.1)

$$
\begin{cases}
\dfrac{\partial \log L\big(\mathbf{x}_{r,n}; \boldsymbol{\theta} | \mathbf{x}_{\cdot 1}\big)}{\partial \mu_k} = 0 \;\Leftrightarrow\; \displaystyle\sum_{t=2}^{n} \frac{P_t(x_{k,t}-1)}{P_t(x_{k,t})} = (n-1) \;, \\
\dfrac{\partial \log L\big(\mathbf{x}_{r,n}; \boldsymbol{\theta} | \mathbf{x}_{\cdot 1}\big)}{\partial \alpha_k} = 0 \;\Leftrightarrow\; \displaystyle\sum_{t=2}^{n} x_{k,t} - \alpha_k \sum_{t=2}^{n} x_{k,t-1} - \mu_k \sum_{t=2}^{n} \frac{P_t(x_{k,t}-1)}{P_t(x_{k,t})} = 0 \;,
\end{cases}
$$

where

$$
P_t(y) \;=\; \exp\big[-(\lambda_k+\delta)\big] \sum_{i=0}^{M_{k,t}} \frac{(\lambda_k+\delta)^{y-i}}{(y-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i} \;.
$$

These equations do not yield explicit forms for the estimators of μ_k and α_k, therefore iterative methods are used to solve the system. We use the bisection method, halving the amplitude of the interval which contains the zero of the function until the required precision is obtained.

3.1.2. Conditional Least Squares Estimators

To obtain the Conditional Least Squares (CLS) estimators, we proceed similarly to Al-Osh and Alzaid (1987) in the analysis of PoINAR(1) model. Thus,

the Conditional Least Squares (CLS) estimator of the parameter is obtained by minimizing

$$(3.2) \quad Q = \sum_{k=1}^{r}\sum_{t=2}^{n}\big[X_{k,t} - \mathrm{E}(X_{k,t}|X_{k,t-1})\big]^2 = \sum_{k=1}^{r}\sum_{t=2}^{n}\big[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big]^2 .$$

Therefore, calculating the derivatives of the previous expression in order to α_k, λ_k and δ, we obtain respectively

$$(3.3) \quad \begin{cases} \partial Q/\partial\alpha_k = -2\sum_{t=2}^{n} X_{k,t-1}\big[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big] , \\ \partial Q/\partial\lambda_k = -2\sum_{t=2}^{n}\big[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big] , \qquad k=1,...,r . \\ \partial Q/\partial\delta = -2\sum_{k=1}^{r}\sum_{t=2}^{n}\big[X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big] , \end{cases}$$

Setting the derivatives to zero, we observe that $\partial Q/\partial\delta$ is a multiple of $\partial Q/\partial\lambda_k$. It is easy to check that the normal equations constitute an indeterminate system and, similarly to the maximum likelihood method, it is not possible to estimate the parameters $\delta, \alpha_k, \lambda_k$, $k=1,...,r$, separately. Therefore, once again we consider $\mu_k = \lambda_k + \delta$ in expression (3.2).

After some simple algebraic operations the estimators are given by

$$\hat{\alpha}_{k,CLS} = \frac{(n-1)\sum_{t=2}^{n} X_{k,t}X_{k,t-1} - \big(\sum_{t=2}^{n} X_{k,t}\big)\big(\sum_{t=2}^{n} X_{k,t-1}\big)}{(n-1)\sum_{t=2}^{n} X_{k,t-1}^2 - \big(\sum_{t=2}^{n} X_{k,t-1}\big)^2} ,$$

$$\hat{\mu}_{k,CLS} = \frac{\sum_{t=2}^{n} X_{k,t} - \hat{\alpha}_{k,LSE}\sum_{t=2}^{n} X_{k,t-1}}{(n-1)} .$$

3.1.3. Moment Estimators

Considering that the one step ahead prediction error is

$$e_{k,t} = X_{k,t} - \mathrm{E}(X_{k,t}|X_{k,t-1}) , \qquad k=1,2,...,r ,$$

we have that $\mathrm{E}(e_{k,t}|X_{k,t-1}) = 0$, $\mathrm{E}(X_{k,t-1}e_{k,t}|X_{k,t-1}) = 0$ and the corresponding sample moments are the following

$$(3.4) \quad \begin{cases} \dfrac{1}{n-1}\sum_{t=2}^{n}\big(X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big) = 0 , \\ \dfrac{1}{n-1}\sum_{t=2}^{n} X_{k,t-1}\big(X_{k,t} - \alpha_k X_{k,t-1} - \lambda_k - \delta\big) = 0 , \end{cases}$$

for $k=1,2,...,r$. This system has $2r$ equations and $2r+1$ unknown parameters so it will be necessary to add another equation in order to estimate all the parameters. Through the analysis of covariance matrix given in (2.1), we observe that

$$\mathrm{Cov}(X_{i,t},X_{j,t})-\frac{\delta}{1-\alpha_i\,\alpha_j}=0\ ,\qquad i,j=1,2,...,r,\quad i\neq j\ ,$$

being the corresponding sample moment given by

$$\frac{1}{n-1}\sum_{t=2}^{n}(X_{i,t}-\bar{X}_{i}.)\,(X_{j,t}-\bar{X}_{j}.)-\frac{\delta}{1-\alpha_i\,\alpha_j}\ ,\qquad i,j=1,2,...,r\ , \tag{3.5}$$

with $\bar{X}_{k}.=\sum_{t=2}^{n}X_{k,t}/(n-1),\ k=1,2,...,r.$

Each of these equations yields an estimator for δ. Averaging the $r(n-1)/2$ equations we obtain the following smoothed estimator for δ

$$\frac{2}{r(r-1)}\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\left[\frac{1}{n-1}\sum_{t=2}^{n}(X_{i,t}-\bar{X}_{i}.)\,(X_{j,t}-\bar{X}_{j}.)-\frac{\delta}{1-\alpha_i\,\alpha_j}\right]. \tag{3.6}$$

Thus, from the system (3.4) and equation (3.6), the following estimators for the parameters δ, α_k and λ_k, $k=1,...,r$, are obtained

$$\hat{\alpha}_{k,MM}\ =\ \frac{(n-1)\,\sum_{t=2}^{n}X_{k,t}X_{k,t-1}-\left(\sum_{t=2}^{n}X_{k,t}\right)\left(\sum_{t=2}^{n}X_{k,t-1}\right)}{(n-1)\,\sum_{t=2}^{n}X_{k,t-1}^{2}-\left(\sum_{t=2}^{n}X_{k,t-1}\right)^{2}}\ ,$$

$$\hat{\delta}_{MM}\ =\ \frac{\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{t=2}^{n}\left[(X_{i,t}-\bar{X}_{i}.)\,(X_{j,t}-\bar{X}_{j}.)\right]}{(n-1)\,\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\left[1/(1-\alpha_i\,\alpha_j)\right]}\ ,$$

$$\hat{\lambda}_{k,MM}\ =\ \frac{\sum_{t=2}^{n}X_{k,t}\ -\ \hat{\alpha}_{k,MM}\sum_{t=2}^{n}X_{k,t-1}}{(n-1)}\ -\ \hat{\delta}_{MM}\ .$$

Note that the following relations may be established:

- for α_k moment estimators are the same as conditional least squares estimators, $\hat{\alpha}_{k,MM}=\hat{\alpha}_{k,CLS}$,
- moment estimators for λ_k may be expressed as $\hat{\lambda}_{k,MM}=\hat{\mu}_{k,CLS}-\hat{\delta}_{MM}$.

3.2. Bayesian Approach

It is well known that Bayesian inference is based on the posterior distribution, since this distribution contains all the available information about the unknown parameters $\boldsymbol{\theta}$. After observing the particular sample $\mathbf{x}_n$, the updated information about $\boldsymbol{\theta}$ is expressed by Bayes theorem through posterior distribution which is given by

$$\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})\ =\ \frac{L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{\cdot 1})\,\pi(\boldsymbol{\theta})}{\displaystyle\int_{\Theta}L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_{\cdot 1})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}\ \propto\ L(\mathbf{x}_{r,n};\boldsymbol{\theta}|\mathbf{x}_1)\,\pi(\boldsymbol{\theta})\ ,\qquad \boldsymbol{\theta}\in\Theta\ , \tag{3.7}$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution. In a Bayesian framework it is necessary to assign priors to each parameter. In this work, the prior distributions considered are the beta and gamma distributions since they are conjugated of binomial and Poisson distributions, respectively. Therefore, beta distribution with parameters $a_k, b_k > 0$ is the prior for α_k, $\alpha_k \frown \mathrm{Be}(a_k, b_k)$, and gamma distributions with parameters $c_k, d_k > 0$, $\lambda_k \frown \mathrm{Ga}(c_k, d_k)$ and $e, f > 0$, $\delta \frown \mathrm{Ga}(e, f)$ are the priors for λ_k and δ, respectively.

Moreover, we assume independence between α_k, λ_k and δ, for $k = 1, 2, ..., r$, as well as the knowledge of hiperparameters a_k, b_k, c_k, d_k, e and f, $k = 1, 2, ..., r$. Therefore, the prior distribution of the $2r+1$ parameters $(\alpha_1, \alpha_2, ..., \alpha_r;\ \lambda_1, \lambda_2, ..., \lambda_r;\ \delta)$ has the form

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= \pi(\delta) \prod_{k=1}^{r} \pi(\alpha_k)\, \pi(\lambda_k) \\ &\propto \delta^{e-1} \exp(-f\delta) \prod_{k=1}^{r} {\alpha_k}^{a_k-1} (1-\alpha_k)^{b_k-1} \lambda_k^{c_k-1} \exp(-d_k\lambda_k)\ . \end{aligned} \tag{3.8}$$

Thus, by Bayes theorem it follows from the prior and the likelihood (3.1), that the posterior distribution is given by the following expression

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{x}_{r,n}) \ \propto\ & \delta^{e-1} \exp(-f\delta) \left(\prod_{k=1}^{r} {\alpha_k}^{a_k-1} (1-\alpha_k)^{b_k-1} \lambda_k^{c_k-1} \exp(-d_k\lambda_k) \right) \times \\ \times & \left(\prod_{k=1}^{r} \prod_{t=2}^{n} \sum_{i=0}^{M_{k,t}} \exp\big[-(\lambda_k+\delta)\big] \frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i} \right). \end{aligned} \tag{3.9}$$

The Bayes estimate for $\boldsymbol{\theta}$ is the mean of this distribution which cannot be obtained analytically. Thus we use the Gibbs sampler in order to generate values of $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$. Through Gibbs sampler and based on a irreducible Markov chain with state space Θ whose stationary distribution is $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$, a sequence of correlated realizations is generated. In this context the algorithm is based on the fact that if the joint distribution $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$ is positive over its entire domain, then it is uniquely determined by the m full conditional distributions $\pi(\theta_i|\mathbf{x}_{r,n}, \boldsymbol{\theta}_{-i})$, $i = 1, 2, ..., m$, where $\boldsymbol{\theta}_{-i}$ represents the vector $\boldsymbol{\theta}$ after being removed θ_i component (Besag, 1974; Gelfand and Smith, 1990).

The full conditional posterior densities are

- for α_k

$$\begin{aligned} \pi\big(\alpha_k|\boldsymbol{\alpha}_{-k}, \boldsymbol{\lambda}, \delta, \mathbf{x}_{r,n}\big) &= \pi\big(\alpha_k|\lambda_k, \delta, \mathbf{x}_{k\cdot}\big) \ \propto \\ &\propto \alpha_k^{a_k-1}(1-\alpha_k)^{b_k-1} \prod_{t=2}^{n} \sum_{i=0}^{M_{k,t}} \frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!} \binom{x_{k,t-1}}{i} \alpha_k^i (1-\alpha_k)^{x_{k,t-1}-i}\,, \end{aligned}$$

with $\boldsymbol{\alpha}_{-k} = (\alpha_1, ..., \alpha_{k-1}, \alpha_{k+1}, ..., \alpha_r)$, $\mathbf{x}_{k\cdot} = (x_{k,t}\colon\ t = 1, 2, ..., n)$;

- for λ_k

$$\pi\big(\lambda_k|\boldsymbol{\lambda}_{-k},\alpha,\delta,\mathbf{x}_{r,n}\big) \;=\; \pi\big(\lambda_k|\alpha_k,\delta,\mathbf{x}_{k\cdot}\big) \;\propto\; \lambda_k^{c_k-1}\exp\big[-(\lambda_k\, d_k)\big] \times$$
$$\times \prod_{t=2}^{n}\sum_{i=0}^{M_{k,t}} \exp\big[-(\lambda_k+\delta)\big]\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}\,,$$

with $\boldsymbol{\lambda}_{-k}=(\lambda_1,...,\lambda_{k-1},\lambda_{k+1},...,\lambda_r)$;

- for δ

$$\pi\big(\delta|\boldsymbol{\alpha},\boldsymbol{\lambda},\mathbf{x}_{r,n}\big) \;\propto\; \delta^{e-1}\exp(-f\delta)\times$$
$$\times \prod_{k=1}^{r}\prod_{t=2}^{n}\sum_{i=0}^{M_{k,t}} \exp\big[-(\lambda_k+\delta)\big]\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}\,.$$

The generation of pseudo-random numbers through the full conditional posterior densities may be achieved through the Adaptive Rejection Sampling (ARS) if the functions were surely log-concave. However, since this is not generally the case, we use Adaptive Rejection Metropolis Sampling (ARMS), which is an hybrid method introduced by Gilks *et al.* (1995). Thus, in Gibbs sampler each value $\boldsymbol{\theta}_{-i}$ is generated from $\pi(\theta_i|\mathbf{x}_{r,n},\boldsymbol{\theta}_{-i})$ through ARMS algorithm in the following way:

Algorithm 1.

1. generate a random sample of the model (1.2);

2. calculate the initial estimates of $\alpha_1,...,\alpha_r$ and δ, by the moments method; denote them by $\alpha_{1,0},...,\alpha_{r,0}$ and δ_0;

3. using ARMS method, simulate for each $k=1,2,...,r$,

$$\text{and}\qquad \begin{aligned}\lambda_{k,1} \;&\text{ from }\; \pi\big(\lambda_k|\mathbf{x}_{k\cdot},\delta_0,\alpha_{k,0}\big)\\ \alpha_{k,1} \;&\text{ from }\; \pi\big(\alpha_k|\mathbf{x}_{k\cdot},\delta_0,\lambda_{k,1}\big)\;;\end{aligned}$$

4. simulate, using ARMS method,

$$\delta_1 \;\text{ from }\; \pi\big(\delta|\mathbf{x}_{r,n},\alpha_{1,1},...,\alpha_{r,1},\lambda_{1,1},...,\lambda_{r,1}\big);$$

5. repeat steps 3. and 4. with $i=2,...,nig$ (number of Gibbs sampler iterations); that is, for $k=1,2,...,r$,

$$\begin{aligned}\lambda_{k,i} \;&\text{ is simulated from }\; \pi\big(\lambda_k|\mathbf{x}_{k\cdot},\delta_{i-1},\alpha_{k,i-1}\big)\;,\\ \alpha_{k,i} \;&\text{ is simulated from }\; \pi\big(\alpha_k|\mathbf{x}_{k\cdot},\delta_{i-1},\lambda_{k,i}\big)\\ \delta_i \;&\text{ is simulated from }\; \pi\big(\delta|\mathbf{x}_{r,n},\alpha_{1,i},...,\alpha_{r,i},\lambda_{1,i},...,\lambda_{r,i}\big);\end{aligned}$$

6. despising the first b values (corresponding to the burn-in period) and picking up each value, obtain a sample with $m=(nig-b)/l$ elements. Denote the corresponding sample means by: $\alpha_{k,B}^{(i)}$, $\lambda_{k,B}^{(i)}$ and $\delta_B^{(i)}$;

7. repeat *nrep* times the steps 1. to 6..

Afterwards Bayes estimates can be calculated through the expressions

$$\hat{\alpha}_{k,B} = \frac{1}{nrep}\sum_{i=1}^{nrep}\alpha_{k,B}^{(i)}, \quad \hat{\lambda}_{k,B} = \frac{1}{nrep}\sum_{i=1}^{nrep}\lambda_{k,B}^{(i)} \quad \text{and} \quad \hat{\delta}_B = \frac{1}{nrep}\sum_{i=1}^{nrep}\delta^{(i)} .$$

4. PREDICTIVE INFERENCE

Let $\mathbf{x}_n = \{X_{k,t}\colon k=1,...,r,\ t=2,...,n\}$ be a sample generated by the Poisson SUINAR(1) model. We aim at obtaining the h-step-ahead predictor of $X_{k,n+h}$, $\hat{X}_{k,n+h}$. We begin by presenting some results fundamental to the understanding of the work.

According to the definition of the SUINAR(1) process, we have that

$$X_{k,n+h} = \alpha_k \circ X_{k,n+h-1} + \epsilon_{k,n+h} . \tag{4.1}$$

Iterating backwards h times, equation (4.1) can be written as

$$X_{k,n+h} = \alpha_k^h \circ X_{k,n} + \sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j} , \qquad h=1,2,... .$$

Since $X_{k,n}$ is independent of $\epsilon_{k,n+j}$, $j=1,...,h$, the conditional distribution of $X_{k,n+h}$ on $X_{k,n}$ is

$$P\Big(X_{k,n+h}=x \,\Big|\, X_{k,n}\Big) = P\left(\alpha_k^h\circ X_{k,n} + \sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j} = x \,\middle|\, X_{k,n}\right) =$$
$$= \sum_{y=0}^{\min X_{k,n},x} P\Big(\alpha_k^h\circ X_{k,n} = y \,\Big|\, X_{k,n}\Big)\, P\left(\sum_{j=1}^{h}\alpha_k^{h-j}\circ\epsilon_{k,n+j} = x-y\right) .$$

Noting that $\alpha_k\circ X_{k,n}\,|\,X_{k,n} \sim Bi(X_{k,n},\alpha_k)$ and $\epsilon_{k,t}\sim P(\lambda_k)$, it follows easily that the distribution of $X_{k,n+h}|X_{k,n}$ is the convolution of the distribution of the innovation process, a Poisson distribution with parameter $(\lambda_k+\delta)(1-\alpha_k^h)/(1-\alpha_k)$, and that resulting from the binomial thinning operation, a binomial distribution with parameters $X_{k,n}$ and α_k^h. This result, proved in Silva (2005), is established in the following theorem:

Theorem 4.1. *For the Poisson SUINAR(1) model, the distribution of $X_{k,n+h}$ given $X_{k,n}$ is the convolution of a binomial distribution with parameters $X_{k,n}$ and α_k^h and a Poisson distribution with parameter $(\lambda_k+\delta)(1-\alpha_k^h)/(1-\alpha_k)$. That is to say, $X_{k,n+h}|X_{k,n}$ has the moment generating function*

$$\varphi_{X_{k,n+h}|X_{k,n}}(s) = \left[\alpha_k^h\, e^s + (1-\alpha_k^h)\right]^{x_{k,n}} \exp\left\{(\lambda_k+\delta)\,\frac{1-\alpha_k^h}{1-\alpha_k}\,(e^s-1)\right\} . \tag{4.2}$$

Thus, the probability function of $X_{k,n+h}|X_{k,n}$, $k=1,2...,r$, is given by

$$\begin{aligned} p(x_{k,n+h}|x_{k,n}) &= P\Big(X_{k,n+h}=x \,\Big|\, X_{k,n}=x_{k,n}\Big) \\ &= \sum_{i=0}^{\min(x,x_{k,n})} \binom{x_{k,n}}{i} (\alpha_k^h)^i \, (1-\alpha_k^h)^{x_{k,n}-i} \;\times \\ &\quad \times \exp\left[-(\lambda_k+\delta)\,\frac{1-\alpha_k^h}{1-\alpha_k}\right] \frac{1}{(x-i)!}\left[(\lambda_k+\delta)\,\frac{1-\alpha_k^h}{1-\alpha_k}\right]^{x-i}, \quad k=1,2,...,r\,. \end{aligned} \tag{4.3}$$

Since $\lim\limits_{h\to+\infty} \varphi_{X_{k,n+h}|X_{k,n}}(s) = \exp\left[\dfrac{\lambda_k+\delta}{1-\alpha_k}\,(e^s-1)\right]$, the corollary follows.

Corollary 4.1. *$X_{k,n+h}|X_{k,n}$ has the Poisson limit distribution with parameter $(\lambda_k+\delta)/(1-\alpha_k)$.*

4.1. Classical Prediction

4.1.1. Forecasts of future observations

Analogously to the study made by Silva *et al.* (2006) concerning prediction in PoINAR(1) processes, we will calculate two predictors of $X_{k,n+h}$. One of them is based on the minimization of mean square error and the other minimizes the mean absolute error. Due to the fact that the best predictor which minimizes the mean square error is $\hat{X}_{k,n+h} = \mathrm{E}[X_{k,n+h}|X_{k,n}]$ and according to expression (4.2), it comes straightforwardly that $\mathrm{E}[X_{k,n+h}|X_{k,n}] = \varphi'_{X_{k,n+h}|X_{k,n}}(s)_{|s=0}$. Therefore

$$\hat{X}_{k,n+h} = \mathrm{E}[X_{k,n+h}|X_{k,n}] = \alpha_k^h X_{k,n} + \frac{1-\alpha_k^h}{1-\alpha_k}\,(\lambda_k+\delta)\;, \qquad k=1,2,...,r\;. \tag{4.4}$$

This method hardly produces coherent predictions in the sense that forecasts of integer values must be integer values as well (see Chatfield, 2001). In order to obtain coherent predictions for X_{n+h}, Freeland and McCabe (2003) suggest using the value which minimizes the expected absolute error given the sample, i.e., the value that minimizes $\mathrm{E}\big[|X_{n+h}-\hat{X}_{n+h}| \,\big|\, X_n\big]$. Let $m_{k,h}$ be the median of the conditional distribution $X_{k,n+h}|X_{k,n}$. It can be proved that $\mathrm{E}\big[|X_{k,n+h}-\hat{m}_{k,n+h}| \,\big|\, X_{k,n}\big]$ has a global minimum in $\hat{m}_{k,n+h}=m_{k,h}$; in this sense, this means that median of the predictive distribution is the best predictor of $X_{k,n+h}$.

4.1.2. Prediction Intervals

A prediction interval is always more informative than a point forecast. The method for obtaining confidence intervals for the predicted value is based on the probability function of the h-steps-ahead forecast error, which is given by

$$e_{k,n+h}|\mathbf{x}_{r,n} = X_{k,n+h} - \hat{X}_{k,n+h} = X_{k,n+h} - \alpha_k^h x_{k,n} - \frac{1-\alpha_k^h}{1-\alpha_k}\,(\lambda_k+\delta)\ .$$

It is worth to mention that $e_{k,n+h}$ is a discrete variable taking values on $\big\{j - \alpha_k^h\, x_{k,n} - \big[(\lambda_k+\delta)\,(1-\alpha_k^h)/(1-\alpha_k)\big];\ j = 0,1,2...\big\}$; hence has the probability function,

$$P\left(e_{k,n+h}|\mathbf{x}_{r,n} = j - \alpha_k^h\, x_{k,n} - (\lambda_k+\delta)\,\frac{1-\alpha_k^h}{1-\alpha_k}\right) = P\Big(X_{k,n+h}=j\,\Big|\,X_{k,n}=x_{k,n}\Big) =$$

$$= \exp\left[-(\lambda_k+\delta)\,\frac{1-\alpha_k^h}{1-\alpha_k}\right] \times$$

$$\times \sum_{i=0}^{\min(j,x_{k,n})} \frac{\big[(\lambda_k+\delta)\,(1-\alpha_k^h)/(1-\alpha_k)\big]^{j-i}}{(j-i)!} \binom{x_{k,n}}{i} (\alpha_k^h)^i\,(1-\alpha_k^h)^{x_{k,n}-i}\ .$$

Once the probability function of the forecast error is known, the $100\gamma\%$ confidence interval for $X_{k,n+h}$ is given by

$$\big(\hat{X}_{k,n+h} + e_{t_1},\, \hat{X}_{k,n+h} + e_{t_2}\big)\ , \tag{4.5}$$

where $\hat{X}_{k,n+h}$ is defined by (4.4), e_{t_1} is the greatest value $e_{k,n+h}|\mathbf{x}_{r,n}$ such as $P(e_{k,n+h}|\mathbf{x}_{r,n} \le e_{t_1}) \le (1-\gamma)/2$ and e_{t_2} is the lowest value of $e_{k,n+h}|\mathbf{x}_{r,n}$, such as $P(e_{k,n+h}|\mathbf{x}_{r,n} \le e_{t_2}) \ge (1+\gamma)/2$.

4.2. Bayesian Prediction

To obtain the Bayesian predictive function we use the randomness of both the future observation $X_{k,n+h}$ we want to predict and the vector of unknown parameters $\boldsymbol{\theta}$. Moreover, information about $\boldsymbol{\theta}$ is contained in the observed sample $\mathbf{x}_{r,n}$ and is quantified on the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$. Thus the following definition.

Definition 4.1. Let $\boldsymbol{\theta} \in \Theta$ be the vector of unknown parameters. The h steps-ahead Bayesian posterior predictive distribution is defined by

$$\pi(x_{k,n+h}|\mathbf{x}_{r,n}) = \int_\Theta \pi(x_{n+h};\boldsymbol{\theta}|\mathbf{x}_{r,n})\,d\boldsymbol{\theta} = \int_\Theta p(x_{k,n+h}|\mathbf{x}_{r,n};\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})\,d\boldsymbol{\theta}\ , \tag{4.6}$$

where $\pi(\boldsymbol{\theta}|\mathbf{x}_{r,n})$ is the posterior probability density function of $\boldsymbol{\theta}$ and $p(x_{k,n+h}|\mathbf{x}_{r,n};\boldsymbol{\theta})$ is the classic predictive function.

The predictive distribution $X_{n+h}|\mathbf{x}_{r,n}$ given by (4.6) is looked upon as containing all the accumulated information on the future values. Therefore, the Bayesian predictor of $X_{k,n+h}$ can be calculated through the mean value, the median or the mode of the predictive function $\pi(x_{k,n+h}|\mathbf{x}_{r,n})$.

4.2.1. Forecasts of future observations

According to Definition 4.1, the h-steps-ahead Bayesian predictive function for the k-th individual of the SUINAR(1) model is given by

$$\begin{aligned}\pi(x_{k,n+h}|\mathbf{x}_{r,n}) &= \int_{\Theta_k} \pi(x_{k,n+h},\boldsymbol{\theta}_k|\mathbf{x}_{r,n})\, d\boldsymbol{\theta}_k \\ &= \int_{\Theta_k} p(x_{k,n+h}|\mathbf{x}_{r,n},\boldsymbol{\theta}_k)\, \pi(\boldsymbol{\theta}_k|\mathbf{x}_{r,n})\, d\boldsymbol{\theta}_k \\ &= \int_{\Theta_k} p(x_{k,n+h}|x_{k,n},\boldsymbol{\theta}_k)\, \pi(\boldsymbol{\theta}_k|\mathbf{x}_{r,n})\, d\boldsymbol{\theta}_k\ ,\end{aligned} \tag{4.7}$$

where $\theta_k=(\delta,\alpha_k,\lambda_k)$, $p(x_{k,n+h}|x_{k,n},\theta_k)$, $k=1,2,...,r$, is given by (4.3) and $\pi(\theta_k|\mathbf{x}_{r,n})$ is the posterior probability density function of θ_k defined by

$$\begin{aligned}\pi(\theta_k|\mathbf{x}_{r,n}) &\propto \pi(\theta_k)\, L(\mathbf{x}_{r,n},\delta,\lambda_k,\alpha_k|\mathbf{x}_{\cdot 1}) \\ &\propto \delta^{e-1}\exp(-f\delta)\ {\alpha_k}^{a_k-1}(1-\alpha_k)^{b_k-1}\,\lambda_k^{c_k-1}\exp(-d_k\lambda_k)\ \times \\ &\quad\times\left(\prod_{t=2}^{n}\sum_{i=0}^{M_{k,t}}\exp\big[-(\lambda_k+\delta)\big]\frac{(\lambda_k+\delta)^{x_{k,t}-i}}{(x_{k,t}-i)!}\binom{x_{k,t-1}}{i}\alpha_k^i(1-\alpha_k)^{x_{k,t-1}-i}\right).\end{aligned}$$

Usually, $X_{k,n+h}$ is predicted by $\mathrm{E}(X_{k,n+h}|\mathbf{x}_{r,n})$ which does not seem feasible here due to the complexity of equation (4.7). Thus we propose two methodologies to deal with the problem. In the first approach, using the expected value properties, $\mathrm{E}(X_{k,n+h}|\mathbf{x}_{r,n})$ is rewritten as follows:

$$\begin{aligned}\mathrm{E}\big[X_{k,n+h}|\mathbf{x}_{r,n}\big] &= \mathrm{E}\Big[\mathrm{E}(X_{k,n+h}|\mathbf{x}_{r,n},\theta_k)\,\Big|\,\mathbf{x}_{r,n}\Big] \\ &= \mathrm{E}\Big[\alpha_k^h X_{k,n}+(1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,\Big|\,\mathbf{x}_{r,n}\Big] \qquad \text{by (4.4)} \\ &= X_{k,n}\,\mathrm{E}(\alpha_k^h|\mathbf{x}_{r,n})+\mathrm{E}\Big[(1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,\Big|\,\mathbf{x}_{r,n}\Big]\ .\end{aligned}$$

Now, the mean values $\mathrm{E}(\alpha_k^h|\mathbf{x}_{r,n})$ and $\mathrm{E}\big[(1-\alpha_k^h)(\lambda_k+\delta)/(1-\alpha_k)\,|\,\mathbf{x}_{r,n}\big]$, can be estimated using Gibbs methodology jointly with ARMS algorithm to generate m values of the full conditional distributions: $\big(\delta^{(1)},\delta^{(2)},...,\delta^{(m)}\big)$, $\big(\alpha_k^{(1)},\alpha_k^{(2)},...,\alpha_k^{(m)}\big)$ and $\big(\lambda_k^{(1)},\lambda_k^{(2)},...,\lambda_k^{(m)}\big)$ for $k=1,2,...,r$, necessary to the evaluation of the corresponding ergodic means (see Section 3.2). Thus, $X_{k,n+h}$ can be estimated by

$$\hat{X}_{k,n+h} = x_{k,n}\,\frac{1}{m}\sum_{i=1}^{m}\big(\alpha_k^{(i)}\big)^h+\left[\frac{1}{m}\sum_{i=1}^{m}\frac{1-\big(\alpha_k^{(i)}\big)^h}{1-\alpha_k^{(i)}}\left(\lambda_k^{(i)}+\delta^{(i)}\right)\right], \tag{4.8}$$

where m is the number of replications really used, after convergency attained.

The second approach applies Tanner composition method, Tanner (1996), to the SUINAR(1) model. A sample $(X_{k,n+h,1}, X_{k,n+h,2}, ..., X_{k,n+h,m})$ is generated from the predictive distribution (4.7) using Algorithm 2 described bellow. Then, the forecast for the future observation $X_{k,n+h}$ can be calculated through the sample mean, median or mode.

Algorithm 2.

1. Calculate an initial estimate α_0 and δ_0 for α_k and δ, respectively, using MM estimation from a sample $\{X_{k,t}: \ k = 1, ..., r, \ t = 2, ..., n\}$ of the Poisson SUINAR(1) defined by (1.2);

2. using Gibbs methodology jointly with adaptive rejection Metropolis sampling (ARMS), sample values of the triplets $(\alpha_{k,1}, \lambda_{k,1}, \delta_1)$, $(\alpha_{k,2}, \lambda_{k,2}, \delta_2)$, ..., $(\alpha_{k,m}, \lambda_{k,m}, \delta_m)$ from the full conditional distributions of α_k, λ_k and δ;

3. for each i $(i = 1, ..., m)$ draw $X_{k,n+h,i}$ from $\pi(x_{k,n+h}|x_{r,n}, \alpha_{k,i}, \lambda_{k,i}, \delta_i)$, using the inverse transformation method adapted to discrete variables. That means:

(**a**) sample a scaler u from Uniform distribution $U(0, 1)$,

(**b**) evaluate the lowest integer value s: $\sum_{i=0}^{s} \pi\big(x_{k,n+h}|x_{r,n}, \alpha_i, \lambda_i, \delta_i\big) \geq u$,

(**c**) consider $X_{k,n+h,i} = s$.

Thus, we have sampled $X_{k,n+h,1}$, $X_{k,n+h,2}$, ..., $X_{k,n+h,m}$ from the posterior predictive distribution.

4.2.2. HPD predictive intervals

In this section Highest Probability Density (HPD) predictive intervals are obtained from the posterior predictive distribution (Paulino *et al.*, 2003).

Definition 4.2. $R(\gamma) = (X_L, X_R)$ is a prediction interval HPD (degree γ) for $X_{k,n+h}$ if

$$P\big(X_L \leq X_{k,n+h} \leq X_R\big) \ = \sum_{x_{k,n+h}=X_L}^{X_R} \pi(x_{k,n+h}|\mathbf{x}_{r,n}) \ \geq \ K_\gamma \ ,$$

where K_γ is the largest constant such that $P\big[X_{n+h} \in R(\gamma)\big] \geq \gamma$.

The computation of the HPD interval for $X_{k,n+h}$ is hindered by the lack of an explicit expression for the posterior predictive probability function, equation (4.7). However an estimate of $R(\gamma)$ may be obtained using Chen and Shao (1999) algorithm which is outlined next.

Algorithm 3.

1. draw a sample from $\pi(x_{k,n+h}|\mathbf{x}_{r,n})$ (Algorithm 2);

2. order the sample values $X_{(k,n+h,1)}$, $X_{(k,n+h,2)}$, ..., $X_{(k,n+h,m)}$, obtained in 1.;

3. for fixed γ, calculate the intervals

$$\hat{R}_i(\gamma) = \left(X_{(k,n+h,i)}, X_{(k,n+h,i+[m\gamma])}\right) , \qquad 1 \leq i \leq m - [m\gamma] ,$$

where $[m\gamma]$ is the integer part of $m\gamma$. Choose for $100\gamma\%$ HPD interval for $X_{k,n+h}$, the $\hat{R}(\gamma)$ with smallest amplitude.

$\hat{R}(\gamma)$ is an estimator of $R(\gamma)$, whose asymptotic properties are valid under certain regularity conditions (Theorem 7.3.1., Chen *et al.*, 2000). Noting that we are considering point processes, the Algorithm 3 can produce more than one interval. When this is the case we choose for $\hat{R}(\gamma)$ the interval with highest absolute frequency, between those with smaller amplitude; in the case of equality of the absolute frequencies, the interval considered is the one with smaller inferior limit as suggested by Chen *et al.* (2000).

5. SIMULATION STUDY

In this section the small sample properties of the estimation and forecasting methods proposed are accessed by means of a simulation study. The data are generated according to model (1.2) with $r = 5$, $\delta = 2$ and for several sets of parameters $(\alpha_1, ..., \alpha_5, \lambda_1, ..., \lambda_5)$. The sets of values for the parameters α_k and λ_k combine small, $\boldsymbol{\alpha}_s$, large, $\boldsymbol{\alpha}_l$ and mixed, $\boldsymbol{\alpha}_{sl}$ values for the α's with small, $\boldsymbol{\lambda}_s$, large $\boldsymbol{\lambda}_l$ and mixed, $\boldsymbol{\lambda}_{sl}$ values for the λ's, in a total of nine models, $(\boldsymbol{\alpha}_s, \boldsymbol{\lambda}_s)$, $(\boldsymbol{\alpha}_s, \boldsymbol{\lambda}_l)$, $(\boldsymbol{\alpha}_s, \boldsymbol{\lambda}_{sl})$, ..., $(\boldsymbol{\alpha}_{sl}, \boldsymbol{\lambda}_{sl})$, described in Table 1. For each model, 200 time series of dimension $n = 25, 50, 100$ are generated.

Table 1: Values of the vector parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ used to simulate the samples.

	α_1	α_2	α_3	α_4	α_5		λ_1	λ_2	λ_3	λ_4	λ_5
$\boldsymbol{\alpha}_s$	0.2	0.2	0.1	0.1	0.2	$\boldsymbol{\lambda}_s$	1.5	1.0	1.0	1.5	1.0
$\boldsymbol{\alpha}_l$	0.8	0.8	0.8	0.9	0.9	$\boldsymbol{\lambda}_l$	3.0	3.0	2.5	2.5	3.0
$\boldsymbol{\alpha}_{sl}$	0.2	0.8	0.9	0.1	0.2	$\boldsymbol{\lambda}_{sl}$	3.0	0.5	1.0	3.0	0.1

5.1. Parameter Estimation

To calculate the Bayesian estimates we use vague prior distributions, considering all the hyperparameters approximately null. This choice is due to the fact that, for one hand we are dealing with simulated samples hence there is no available prior information, and for the other hand the main purpose is to compare the performance between classical and Bayesian methodologies. In Algorithm 1, we set $nig = 3100$, with $b = 1100$ as burn-in period and $l = 20$, to reduce autocorrelation between MCMC samples. A problem that occurs frequently when estimating INAR models by classic methodology is that the estimates for the parameters α_k are inadmissible, that is to say that $\alpha_k \notin (0, 1)$. In this study these samples are eliminated.

The performance of the estimation methods is illustrated in Tables 2 and 3 for two particular situations of the Poisson SUINAR(1) model and based on 200 independent replicates. In Table 2 we consider the model $(\boldsymbol{\alpha}_{sl}, \boldsymbol{\lambda}_{sl})$ with parameters $\boldsymbol{\alpha}_{sl}$: $\alpha_1 = 0.2$, $\alpha_2 = 0.8$, $\alpha_3 = 0.9$, $\alpha_4 = 0.1$, $\alpha_5 = 0.2$, $\boldsymbol{\lambda}_{sl}$: $\lambda_1 = 3.0$, $\lambda_2 = 0.5$, $\lambda_3 = 1.0$, $\lambda_4 = 3.0$, $\lambda_5 = 0.1$ and $\delta = 2$ which is caracterized by both α_k and λ_k ranging from low to high values, meaning that the mean of the innovations varies among the individuals. Table 3 presents the estimation results for the model $(\boldsymbol{\alpha}_s, \boldsymbol{\lambda}_l)$ with parameters $\boldsymbol{\alpha}_s$: $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\alpha_3 = 0.1$, $\alpha_4 = 0.1$, $\alpha_5 = 0.2$, $\boldsymbol{\lambda}_l$: $\lambda_1 = 3.0$, $\lambda_2 = 3.0$, $\lambda_3 = 2.5$, $\lambda_4 = 2.5$, $\lambda_5 = 3.0$ and $\delta = 2$ which is caracterized by low values for the parameters α and high values for the innnovations for all the individuals, with small variation between individuals. These results indicate that the method of moments (mm) provides better estimates for small values of α_k ($\alpha_k \leq 0.2$) whereas the maximum likelihood (ml) and Bayesian methodology (B) are more appropriate when the α_k parameter has large values ($\alpha_k \geq 0.8$); however, the Bayesian approach has the advantage of estimating δ, α_k and λ_k separately, which is not possible with the maximum likelihood. Regarding the estimation of λ_k the simulation results indicate that the Bayesian methodology has a better performance when the mean value of entrances is very different from individual to individual. However, if the differences between the mean values are small, the behavior is not so good. It can be noticed that the method of moments provides always poor estimates for λ_k. Moreover, the parameter δ is underestimated by both methods and the bias increases in the samples where the mean number of entrances differ between the individuals. Regarding the estimation of $\mu_k = \delta + \lambda_k$ the method of moments provides the estimates with smallest bias, whereas the maximum likelihood estimates are the most biased. It is important to note once again that μ_k is estimated as a parameter by ml while $\hat{\mu}_{k,mm} = \hat{\delta}_{mm} + \hat{\lambda}_{k,mm}$.

Table 2: Estimates of $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta)$ model with parameters $\boldsymbol{\alpha}_{sl} = (0.2, 0.8, 0.9, 0.1, 0.2)$, $\boldsymbol{\lambda}_{sl} = (3.0, 0.5, 1.0, 3.0, 0.1)$ and $\delta = 2$ (variances in brackets).

		$n = 25$			$n = 100$		
k	α_k	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$
1	0.2	0.230 (0.02)	0.334 (0.03)	0.256 (0.02)	0.183 (0.89)	0.250 (0.01)	0.197 (0.01)
2	0.8	0.673 (0.02)	0.847 (0.00)	0.842 (0.00)	0.766 (0.01)	0.865 (0.00)	0.865 (0.00)
3	0.9	0.794 (0.02)	0.919 (0.00)	0.918 (0.00)	0.873 (0.00)	0.924 (0.00)	0.924 (0.00)
4	0.1	0.177 (0.02)	0.275 (0.03)	0.224 (0.02)	0.118 (0.01)	0.174 (0.01)	0.125 (0.01)
5	0.2	0.143 (0.02)	0.673 (0.05)	0.623 (0.02)	0.155 (0.01)	0.761 (0.00)	0.758 (0.00)

k	λ_k	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$
1	3.0	3.783 (0.70)		3.333 (0.82)	4.004 (0.29)		3.812 (0.37)
2	0.5	2.445 (2.00)		0.685 (0.13)	1.659 (0.32)		0.782 (0.03)
3	1.0	4.322 (10.86)		1.088 (0.39)	2.494 (0.92)		1.303 (0.07)
4	3.0	3.548 (0.53)		3.016 (0.65)	3.845 (0.14)		3.695 (0.16)
5	0.1	1.082 (0.14)		0.155 (0.00)	1.114 (0.05)		0.154 (0.00)

k	μ_k	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$
1	2.2	3.917 (0.59)	3.381 (0.879)	3.779 (0.08)	4.076 (0.25)	3.739 (0.34)	4.005 (0.37)
2	2.8	2.581 (1.74)	1.097 (0.13)	1.130 (0.07)	1.730 (0.27)	0.977 (0.02)	0.975 (0.03)
3	2.9	4.457 (10.27)	1.524 (0.31)	1.535 (0.07)	2.565 (0.89)	1.488 (0.82)	1.496 (0.07)
4	2.1	3.683 (0.43)	3.241 (0.71)	3.462 (0.08)	3.916 (0.12)	3.667 (0.23)	3.889 (0.16)
5	2.2	1.217 (0.03)	0.436 (0.06)	0.601 (0.08)	1.185 (0.01)	0.331 (0.01)	0.347 (0.00)

	$\hat{\delta}_{mm}$		$\hat{\delta}_B$	$\hat{\delta}_{mm}$		$\hat{\delta}_B$
$\delta = 2$	0.135 (0.11)		0.446 (0.07)	0.071 (0.06)		0.193 (0.04)

Table 3: Estimates of $(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \delta)$ of SUINAR(1) model with parameters $\boldsymbol{\alpha}_s = (0.2, 0.2, 0.1, 0.1, 0.2)$, $\boldsymbol{\lambda}_l = (3.0, 3.0, 2.5, 2.5, 3.0)$ and $\delta = 2$ (variances in brackets).

		$n = 25$			$n = 100$		
k	α_k	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$
1	0.2	0.212 (0.02)	0.323 (0.03)	0.238 (0.01)	0.181 (0.01)	0.243 (0.01)	0.201 (0.01)
2	0.2	0.217 (0.02)	0.320 (0.03)	0.243 (0.01)	0.196 (0.01)	0.267 (0.01)	0.215 (0.01)
3	0.1	0.180 (0.02)	0.306 (0.04)	0.162 (0.01)	0.125 (0.01)	0.204 (0.02)	0.094 (0.01)
4	0.1	0.183 (0.02)	0.310 (0.04)	0.167 (0.01)	0.119 (0.01)	0.189 (0.01)	0.088 (0.00)
5	0.2	0.215 (0.02)	0.325 (0.04)	0.237 (0.01)	0.187 (0.01)	0.256 (0.01)	0.211 (0.01)

k	λ_k	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$	$\hat{\lambda}_{k,mm}$		$\hat{\lambda}_{k,B}$
1	3.0	3.926 (0.70)		1.018 (0.21)	4.079 (0.29)		0.673 (0.18)
2	3.0	3.948 (0.72)		1.050 (0.27)	3.999 (0.24)		0.603 (0.19)
3	2.5	3.174 (0.43)		0.553 (0.10)	3.411 (0.15)		0.255 (0.02)
4	2.5	3.150 (0.38)		0.526 (0.05)	3.432 (0.12)		0.272 (0.03)
5	3.0	3.909 (0.45)		1.026 (0.22)	4.096 (0.22)		0.669 (0.18)

k	μ_k	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$	$\hat{\mu}_{k,mm}$	$\hat{\mu}_{k,ml}$	$\hat{\mu}_{k,B}$
1	2.2	3.947 (0.61)	3.388 (0.89)	3.787 (0.40)	4.079 (0.27)	3.767 (0.38)	3.966 (0.27)
2	2.2	3.968 (0.65)	3.447 (1.00)	3.819 (0.47)	3.999 (0.22)	3.646 (0.33)	3.897 (0.29)
3	2.1	3.195 (0.36)	2.702 (0.66)	3.322 (0.29)	3.412 (0.13)	3.102 (0.27)	3.549 (0.12)
4	2.1	3.171 (0.35)	2.675 (0.65)	3.294 (0.25)	3.433 (0.10)	3.159 (0.20)	3.566 (0.13)
5	2.2	3.930 (0.60)	3.376 (0.98)	3.795 (0.41)	4.096 (0.21)	3.747 (0.32)	3.963 (0.28)

	$\hat{\delta}_{mm}$		$\hat{\delta}_B$	$\hat{\delta}_{mm}$		$\hat{\delta}_B$
$\delta = 2$	0.0211 (0.05)		1.3832 (0.19)	0.0005 (0.01)		1.1759 (0.09)

5.2. Prediction

In this section h-steps-ahead ($h = 1, 2, ..., 10$) point forecasts and prediction intervals are obtained using classic methodology, equations (4.4) and (4.5) and Bayesian methodology, equation (4.8) and Algorithm 3 to obtain HPD predictive intervals. The performance of the forecasting methods is illustrated in Tables 4 and 5 for two particular Poisson SUINAR(1) models.

Table 4: Forecasts for $x_{k,n+h}$ and values of square deviances ($DA^2 = (\hat{x}_{k,n+h} - x_{k,n+h})^2$) of SUINAR(1) model with initial values $\boldsymbol{\alpha}_s = (0.2, 0.2, 0.1, 0.1, 0.2)$, $\boldsymbol{\lambda}_s = (1.5, 1.0, 1.0, 1.5, 1.0)$ and $\delta = 2$.

		$n = 25$					$n = 100$				
h	k	jump	classical		bayesian		jump	classical		bayesian	
			$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2		$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2
1	1	1	2.672	0.107	2.323	0.458	2	3.191	1.418	3.254	1.571
	2	0	2.272	0.530	2.744	0.066	3	2.864	3.474	3.180	4.752
	3	0	2.618	0.146	2.712	0.083	1	2.328	0.452	2.233	0.588
	4	0	2.789	0.045	2.841	0.025	1	3.095	3.629	3.178	3.320
	5	3	1.340	7.076	1.307	7.252	0	2.721	0.078	2.812	0.035
2	1	1	2.857	3.448	2.626	2.644	2	3.135	1.288	3.216	1.479
	2	2	2.213	1.471	2.679	2.819	2	2.514	0.264	2.876	0.767
	3	1	2.528	0.279	2.641	0.411	1	2.359	1.847	2.324	1.753
	4	0	2.697	0.092	2.811	0.036	1	2.956	0.002	3.097	0.009
	5	1	1.523	0.228	1.539	0.213	1	2.617	0.381	2.743	0.552
4	1	2	2.922	1.162	2.933	1.138	1	3.131	0.017	3.200	0.040
	2	1	2.208	3.211	2.552	2.097	1	2.373	0.393	2.575	0.181
	3	1	2.502	0.252	2.697	0.486	2	2.362	2.683	2.460	2.372
	4	3	2.639	11.296	2.811	10.170	1	2.932	0.004	2.991	0.000
	5	2	1.675	1.756	1.812	1.411	1	2.564	2.062	2.721	1.636
8	1	5	2.927	16.589	3.247	14.085	3	3.131	4.541	3.173	4.722
	2	0	2.208	0.627	2.472	0.279	3	2.358	1.844	2.509	2.277
	3	1	2.500	0.250	2.781	0.610	0	2.363	0.132	2.443	0.196
	4	2	2.626	5.636	2.799	4.844	0	2.931	1.143	3.021	0.958
	5	0	1.731	0.534	2.114	1.241	1	2.556	0.309	2.705	0.497
10	1	0	2.927	0.859	3.259	1.585	3	3.131	4.541	3.200	4.840
	2	2	2.208	7.795	2.626	5.636	0	2.358	0.696	2.429	2.468
	3	0	2.500	0.250	2.755	0.060	1	2.363	0.406	2.452	0.300
	4	1	2.625	1.891	3.108	0.796	2	2.931	0.867	3.047	1.096
	5	0	1.735	0.540	2.157	1.339	1	2.556	2.085	2.731	1.610
∞	1		2.927					3.131			
	2		2.208					2.358			
	3		2.500					2.363			
	4		2.623					2.931			
	5		1.736					2.556			

Table 5: Forecasts for $x_{k,n+h}$ and values of square deviances ($DA^2 = (\hat{x}_{k,n+h} - x_{k,n+h})^2$) of SUINAR(1) model with initial values $\boldsymbol{\alpha}_l = (0.8, 0.8, 0.8, 0.9, 0.9)$, $\boldsymbol{\lambda}_l = (3.0, 3.0, 2.5, 2.5, 3.0)$ and $\delta = 2$.

		$n = 25$					$n = 100$				
h	k	jump	classical		bayesian		jump	classical		bayesian	
			$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2		$\hat{x}_{k,n+h}$	DA^2	$\hat{x}_{k,n+h}$	DA^2
1	1	2	20.585	2.002	20.173	3.338	1	23.667	5.443	24.375	2.641
	2	5	24.906	37.137	24.969	36.373	0	18.680	0.102	18.755	0.060
	3	1	14.379	0.386	14.244	0.572	1	19.572	2.039	19.782	1.848
	4	0	38.856	0.733	37.932	0.005	1	24.611	0.151	24.677	0.104
	5	3	35.765	10.465	36.073	8.567	5	37.463	29.844	37.173	26.760
2	1	1	20.992	3.968	20.254	1.573	1	22.811	1.414	23.876	0.015
	2	1	24.025	8.851	24.442	6.543	3	18.442	5.963	18.576	6.636
	3	2	14.876	1.263	14.449	2.253	1	19.273	2.983	19.591	1.985
	4	7	39.266	32.879	37.863	50.937	3	25.156	3.400	25.227	3.144
	5	6	35.765	38.875	36.307	32.410	7	57.844	61.528	37.341	53.890
4	1	1	21.476	0.227	20.686	0.099	0	21.908	9.560	23.224	3.154
	2	3	22.745	0.065	23.153	0.023	1	18.131	0.017	18.406	0.165
	3	1	15.035	4.141	14.895	3.591	2	18.918	9.499	19.326	7.150
	4	5	39.557	11.854	37.774	27.311	6	26.076	15.398	26.395	12.996
	5	9	35.109	97.832	36.589	70.745	3	38.418	19.519	37.582	12.831
8	1	1	21.826	0.682	21.049	0.002	2	21.381	31.573	22.387	21.280
	2	8	21.375	11.391	22.538	20.593	2	17.862	9.847	17.928	9.437
	3	2	15.359	11.283	15.537	12.510	2	18.660	11.156	18.837	10.005
	4	4	39.640	5.570	37.370	21.437	5	27.391	2.589	27.647	1.758
	5	5	34.336	44.409	37.207	14.387	4	39.072	3.717	37.995	9.030
10	1	2	21.882	0.014	20.714	1.654	6	21.317	5.368	22.037	9.223
	2	7	21.026	4.105	21.638	6.959	7	17.808	67.109	18.054	63.139
	3	2	15.425	11.731	15.334	11.116	0	18.619	1.907	18.840	1.346
	4	3	39.643	1.841	37.658	11.169	3	27.855	0.731	28.230	1.513
	5	6	33.989	64.176	37.444	20.757	14	39.252	138.016	38.340	161.188
∞	1		21.937					21.273			
	2		20.384					17.738			
	3		15.510					18.577			
	4		39.645					29.664			
	5		29.221					39.640			

Table 4 displays forecasts for $x_{k,n+h}$, the jump between $x_{k,n}$ and $x_{k,n+h}$, and the squared errors between $\hat{x}_{k,n+h}$ and $x_{k,n+h}$, considering samples of sizes $n = 25$ and $n = 100$ simulated from the model with parameters ($\boldsymbol{\alpha}_s$: $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\alpha_3 = 0.1$, $\alpha_4 = 0.1$, $\alpha_5 = 0.2$), ($\boldsymbol{\lambda}_s$: $\lambda_1 = 1.5$, $\lambda_2 = 1.0$, $\lambda_3 = 1.0$, $\lambda_4 = 1.5$, $\lambda_5 = 1.0$) and $\delta = 2$. Table 5 presents similar results for samples generated from the model with parameters ($\boldsymbol{\alpha}_l$: $\alpha_1 = 0.8$, $\alpha_2 = 0.8$, $\alpha_3 = 0.8$, $\alpha_4 = 0.9$, $\alpha_5 = 0.9$), ($\boldsymbol{\lambda}_l$: $\lambda_1 = 3.0$, $\lambda_2 = 3.0$, $\lambda_3 = 2.5$, $\lambda_4 = 2.5$, $\lambda_5 = 3.0$) and $\delta = 2$.

Additionally Figure 1 presents absolute errors between predicted values and corresponding simulated values, regarding several samples of size 25 of SUINAR(1) model. According to the present simulation study we can conclude that the results are independent of the prediction method and the methodology. Moreover, the observed prediction error depends on two factors: the jump between $x_{k,n}$ and $x_{k,n+h}$ for $h \leq 4$ and the proximity between $x_{k,n+h}$ and $(\hat{\lambda}_k + \hat{\delta})/(1-\hat{\alpha_k})$ for large values of h $(h \geq 5)$ (remark that $\lim_{h\to\infty} \mathrm{E}(X_{k,n+h}|X_{k,n}) = (\lambda_k + \delta)/(1-\alpha_k)$) (see Figure 1). Several simulated examples indicate that the variability of the predictive function increases with the magnitude of α_k and λ_k, justifying that the predictions shown in Table 5 are worst than those in Table 4. Moreover it is worthwhile to mention that the values of $\hat{x}_{k,n+h}$ are constant for $h \geq 8$ (Table 4) when α_k and λ_k are small. In contrast, these values are not constant when α_k and λ_k are large. There is evidence that the confidence interval gets wider as h increases, as expected and converges to the asymptotic interval. However, the rate of convergence is higher for smaller values of α_k and λ_k.

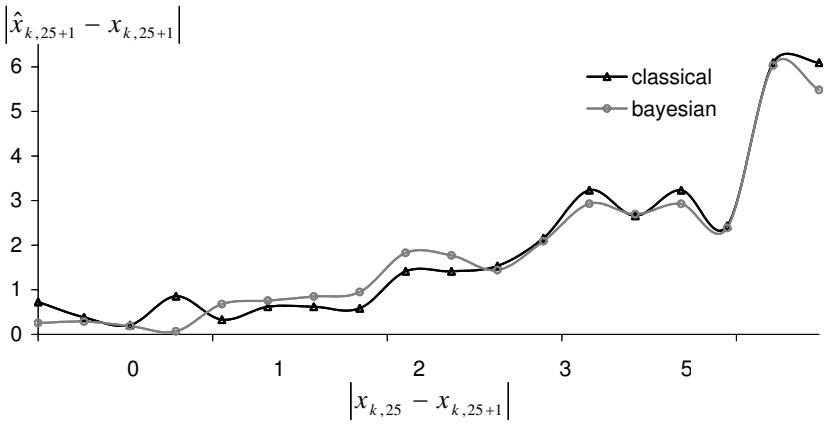

Figure 1: Values of $|\hat{x}_{k,25+1} - x_{k,25+1}|$ with different samples of SUINAR(1) model.

6. APPLICATION

In this section the SUINAR(1) process is used to model the annual number of plants in an industrial sector (electricity, gas, heating and waterpower) in fifteen Swedish municipalities for the period 1984–1993, Berglund and Brännäs (1996). For this data set k is equal to 15 and n is equal to 10. The estimates for the parameters are given in Table 6. From the table it is easily seen that maximum likelihood and Bayes methodologies yield similar estimates only for $k = 3, 5, 8, 10, 11$ and 13. This is due to the small number of observations per individual. In fact, a simulation study with $k = 15$ and $n = 10$ was carried out and it was observed that the three estimation methods yield different estimates and that the differences are larger for small values of α_k and λ_k.

Table 6: Estimated model for the number of plants in electricity, gas, heating and waterpower.

k	$\hat{\alpha}_{k,mm}$	$\hat{\alpha}_{k,ml}$	$\hat{\alpha}_{k,B}$	$\hat{\mu}_{k,mm}$	$\mu\alpha_{k,ml}$	$\mu\alpha_{k,B}$
1	0.667	0.104	0.253	2.667	4.729	4.052
2	0.494	0.787	0.486	3.251	1.753	3.445
3	0.579	0.747	0.628	30.922	18.775	28.106
4	0.037	0.774	0.551	8.136	2.000	3.966
5	0.231	0.758	0.724	14.077	5.120	5.631
6	0.494	0.787	0.483	3.251	1.753	3.463
7	0.167	0.265	0.410	8.444	7.529	6.060
8	0.267	0.382	0.384	6.616	5.543	5.439
9	0.500	0.114	0.351	4.722	7.424	5.693
10	0.331	0.718	0.610	16.694	7.479	9.941
11	0.373	0.785	0.809	11.697	5.114	4.673
12	0.370	0.774	0.555	8.136	2.000	3.937
13	0.261	0.664	0.573	7.663	3.551	4.550
14	0.524	0.787	0.510	3.272	1.404	3.581
15	0.387	0.442	0.192	2.372	2.110	3.444

For illustrative purposes, h steps ahead predictions were obtained, for $h = 1, 2, 3, 4, 5$. The predictions for $h = 1$, 2 and 5 are given in Table 7 and 1 step-ahead predictions for the 15 municipalities are represented in Figure 2. Although the estimates of the model parameters differ, the forecasts obtained by the different methodologies are quite similar.

Table 7: Forecasts for $h = 1, 2, 5$ steps-ahead for the number of plants in electricity, gas, heating and waterpower.

k	$h = 1$		$h = 2$		$h = 5$	
	Classical	Bayes	Classical	Bayes	Classical	Bayes
1	12.67	7.32	11.11	6.72	8.92	6.30
2	6.71	6.81	6.57	7.23	6.45	8.12
3	69.16	68.69	70.98	75.21	73.01	88.12
4	8.40	8.03	8.45	8.56	8.45	9.76
5	18.00	18.05	18.23	19.64	18.30	21.78
6	6.71	6.85	6.57	7.19	6.45	7.72
7	11.44	13.12	10.35	12.71	10.13	12.60
8	10.89	11.54	9.52	10.33	9.04	10.36
9	13.72	11.32	11.58	11.03	9.71	10.53
10	27.94	29.92	25.93	32.33	24.98	33.38
11	19.53	21.66	18.98	22.95	18.68	27.72
12	8.40	7.89	8.45	8.60	8.45	10.49
13	10.28	10.32	10.35	11.01	10.37	12.40
14	6.94	7.07	6.91	7.50	6.88	8.04
15	3.15	4.01	3.59	4.19	3.85	4.44

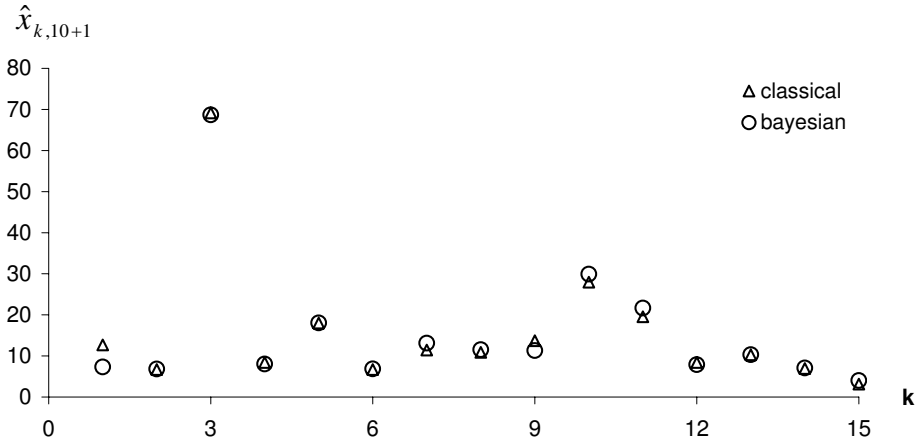

Figure 2: Values of $\hat{x}_{k,10+1}$ for the number of plants relatively to electricity, gas, heating and waterpower.

7. FINAL COMMENTS

In this work classical and Bayesian approaches to time series analysis and forecasting are applied to the SUINAR(1) models. Regarding the estimation of the model, the Bayesian approach has the advantage of allowing the estimation of all the parameters of the model. However, the two methodologies perform similarly regarding the forecasting of future values.

ACKNOWLEDGMENTS

Research (partially) suported by Unidade de Investigação Matemática e Aplicações through Programa Operacional "Ciência, Tecnologia e Inovação" of FCT cofinanced by the European Community fund FEDER. The authors would like to express their gratitude to the Associate Editor and to an anonymous referee for all the helpfull comments.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *J. Time Ser. Anal.*, **8**, 261–275.

[2] Berglund, E. and Brännäs, K. (1996). *Entry and exit of plants: A study based on Swedish panel count data*. In "1995 Yearbook of the Finnish Statistical Society", Helsinki, 95–111.

[3] Berglund, E. and Brännäs, K. (2001). *Plant's entry and exit in Swedish municipalities*. In "The Annals of Regional Science", Springer, Vol. 35, 431–448.

[4] Besag, J. (1974). Spacial interaction and statistical analysis of lattice systems (with discussion), *J. R. Stat. Soc. Ser. B*, **36**, 192–326.

[5] Blundell, R.; Griffith, R. and Windmeijer, F. (1999). Individual effects and dynamics in count data models, *Working Paper*, W99/3, Institute of Fiscal Studies, London.

[6] Böckenholt, U. (1999) Mixed INAR(1) Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data, *J. Econometrics*, **89**, 317–338.

[7] Chatfield, C. (2001). *Time Series Forecasting*, Chapman and Hall/CRC, London.

[8] Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals, *J. Comput. Graph. Statist.*, **8**, 69–92.

[9] Chen, M.-H.; Shao, Q.-M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer Series in Statistics.

[10] Freeland, R.K. and McCabe, B.P.M. (2003). Forecasting discrete valued low count series, *Int. J. Forecasting*, **20**, 427–434.

[11] Gelfand, A. and Smith, A. (1990). Sampling-based approaches to calculating marginal densities, *J. R. Stat. Soc. Ser. B*, **85**, 398–409.

[12] Gilks, W.R.; Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Appl. Statist.*, **44**, 455–472.

[13] McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resour. Bull.*, **21**, 645–650.

[14] Latour, A. (1997). The multivariate GINAR(p) process, *Adv. Appl. Prob.*, **29**, 228–248.

[15] Paulino, C.D.; Amaral Turkman, M.A. and Murteira, B. (2003). *Estatística Bayesiana*, Fundao Calouste Gulbenkian, Lisboa.

[16] Silva, N. (2005). *Análise Bayesiana de séries temporais de valores inteiros*, unpublished PhD thesis, Department of Mathematics, University of Aveiro.

[17] Silva, I.; Silva, M.E.; Pereira, I. and Silva, N. (2005). Replicated INAR(1) process, *Meth. Comp. App. Prob.*, **7**, 517–542.

[18] Silva, N.; Pereira, I. and Silva, M.E. (2006). Forecasting in INAR(1) models, *Cadernos de Matemática*, **CM06;I-46**, Department of Mathematics, University of Aveiro.

[19] Tanner, M.A. (1996). *Tools for Statistical Inference*, 3rd Ed., Springer Verlag, New York.

REVSTAT – Statistical Journal
Volume 6, Number 3, November 2008, 279–307

OPTIMAL AND QUASI-OPTIMAL DESIGNS *

Authors: João Paulo Martins
– Instituto Politécnico de Leiria, Esc. Sup. de Tecnologia e Gestão,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
jpmartins@estg.ipleiria.pt

Sandra Mendonça
– Universidade da Madeira, Dept. de Matemática e Engenharias,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
smfm@uma.pt

Dinis Duarte Pestana
– Universidade de Lisboa, Faculdade de Ciências, DEIO,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
dinis.pestana@fc.ul.pt

Received: July 2008 Revised: October 2008 Accepted: October 2008

Abstract:

- Optimal design theory deals with the choice of the allocation of the observations to accomplish the estimation of some linear combination of the coefficients in a regression model in an optimal way. Canonical moments provide an elegant framework to the theory of optimal designs. An optimal design for polynomial regression of a given degree r can be fatally inappropriate in case the polynomial degree should in fact be s, and hence when r is unknown it would be preferable to consider designs that show good performance for different values of the polynomial degree. Anderson's (1962) path-breaking solution of this multidecision problem has originated many developments, as optimal discriminant designs and optimal robust designs. But once again a design devised for a specific task can be grossly inefficient for a slightly different purpose. We introduce mixed designs; tables for regression of degrees $r = 2,3,4$ exhibiting the loss of efficiency when the optimal mixed design is used instead of the optimal discriminant or of the optimal robust design show that the loss of efficiency is at most 1% and 2%, respectively, while the loss of efficiency when using a discriminant design instead of a robust design or vice-versa can be as high as 10%. Using recursive relations we compute pseudo-canonical moments for measures with infinite support, showing that such pseudo-canonical moments do not share the good identifiability properties of canonical moments of measures whose support is a subset of a compact interval of the real line.

Key-Words:

- *Optimal designs; discriminant designs; robust designs; mixed designs; quasi-optimal designs; canonical and pseudo-canonical moments.*

AMS Subject Classification:

- 62J02, 62K05.

*Invited lecture presented at the IWMS'08 — 17th International Workshop on Matrices and Statistics, in honour of T.W. Anderson's 90th birthday.

1. INTRODUCTION

Suppose the least squares method is used to estimate some linear combination of the coefficients in a regression model $Y(x) = \theta_0 + \theta_1 x + ... + \theta_r x^r + \varepsilon$ on an interval (a, b). The optimal design theory deals with the choice of the allocation of the observations to accomplish the estimation in an optimal way.

The problem has been solved by Smith (1918) using a global optimality criterion based on the variance of the estimated regression function, and *circa* 1960 Guest (1958), Hoel (1958), Box and Draper (1959, 1963), Kiefer (1959, 1961, 1962), Kiefer and Wolfowitz (1959) brought in many new results, namely by introducing sensible optimality criteria, and Anderson (1962) and Kussmaul (1969) investigated the choice of the degree in polynomial regression. See also Stigler (1971) and references therein for the discussion of alternative optimal criteria.

The *design space* $\mathcal{X}$ is the set of all possible points where measurements Y can be taken; $\mathcal{X}$ is assumed to be a compact subset of an Euclidean space. The measurements $Y = Y(x)$, the response at $x \in \mathcal{X}$, is the sum of the deterministic mean effect $\boldsymbol{f}(x)^T\theta = \mathbb{E}[Y|_x]$ and an additive error term ε. In other words,

$$\boldsymbol{Y} = \boldsymbol{f}(x)^T\boldsymbol{\theta} + \varepsilon$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)^T$ is a vector of unknown parameters, $\boldsymbol{f}(x) = \big(f_1(x), ..., f_k(x)\big)^T$ is a vector of real-valued linearly independent continuous regression functions, and ε is an error term with $\mathbb{E}(\varepsilon) = 0$.

For point estimation the *moment assumptions* $\mathbb{E}[\boldsymbol{Y}|_x] = \boldsymbol{f}(x)^T\boldsymbol{\theta}$ and $\mathrm{var}[\boldsymbol{Y}|_x] = \sigma^2 > 0$ provide an adequate setting, but for intervalar estimation or hypothesis testing the usual assumption is that $Y \frown \mathit{Gaussian}\big(\boldsymbol{f}(x)^T\boldsymbol{\theta}, \sigma^2\big)$.

We further assume that the experimenter can take n uncorrelated observations at experimental conditions $x_1, ..., x_n \in \mathcal{X}$

$$Y_i = \boldsymbol{f}(x_i)\,\boldsymbol{\theta} + \varepsilon_i\,, \qquad \mathbb{E}[\varepsilon_i\,\varepsilon_j] = \sigma^2\delta_{ij}\,, \qquad \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \qquad i, j = 1, ..., n$$

at not necessarily distinct points x_i.

Denoting the vectors of the responses $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ and of the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T$, we can rewrite the univariate regression model in matrix form

$$\boldsymbol{Y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

where $X = \big(\boldsymbol{f}(x_1), ..., \boldsymbol{f}(x_n)\big)^T$ denotes the design matrix, $\mathbb{E}(\boldsymbol{Y}) = X\boldsymbol{\theta}$ and the dispersion matrix of the random vector $\boldsymbol{Y}$ is $D(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}_n$.

The estimation of the unknown parameters $\boldsymbol{\theta}, \sigma^2$ from the observed responses $\boldsymbol{Y}$ is an important problem. We shall consider only linear unbiased

estimators $\tilde{\boldsymbol{\theta}}_L = L\boldsymbol{Y}$ where L is a given $k \times n$ matrix and $\mathbb{E}[\tilde{\boldsymbol{\theta}}_L] = LX\boldsymbol{\theta} = \boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \mathbb{R}^k$.

In this general setting, the comparison of linear unbiased estimators is performed in terms of the Loewner ordering of the set of symmetric matrices

$$A \geq B \quad \text{iff} \quad A - B \text{ is nonnegative definite};$$
$$A > B \quad \text{iff} \quad A - B \text{ is positive definite}.$$

It is easily proven (Dette and Studden, 1997, p. 131) that the Gauss–Markov estimator $\tilde{\boldsymbol{\theta}}^{GM} = (X^T X)^{-1} X^T \boldsymbol{Y}$ is BLUE with respect to the Loewner ordering for the regression model with moment assumptions.

Often we are interested in inference about a particular linear combination $z_j\boldsymbol{\theta}$, $z_j \in \mathbb{R}^k$, $j = 1, ..., s$, of the unknown parameters. The *parameter subsystem* $K^T\boldsymbol{\theta} \in \mathbb{R}^s$, where $K = (z_1, ..., z_s)$ denotes a $k \times s$ matrix of rank $s \leq k$ is estimable if and only if there exists a linear unbiased estimator for $K^T\boldsymbol{\theta}$.

This is so if and only if the range inclusion $\text{range}(K) \subseteq \text{range}(X^T)$ is satisfied. In that case, the BLUE for the parameter subsystem $K^T\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}^T = K^T (X^T X)^- X^T \boldsymbol{Y}$$

with minimum dispersion matrix $D(\hat{\boldsymbol{\theta}}^T) = \sigma^2 K^T (X^T X)^- K$.

In the above expression, $(X^T X)^-$ denotes a generalized inverse of $(X^T X)$, i.e. $(X^T X)(X^T X)^- (X^T X) = X^T X$; under the range inclusion condition neither $\hat{\boldsymbol{\theta}}^T$ nor $D(\hat{\boldsymbol{\theta}}^T)$ depend on the specific choice of the generalized inverse.

Under the linear model with gaussian assumption, $K \in \mathbb{R}^{k \times s}$ a given matrix of rank $s \leq k$, if the range inclusion assumption is satisfied for a parameter subsystem $K^T\theta$ and if $n > \text{rank}(X)$, the null hypothesis $H_0 \colon K^T\boldsymbol{\theta} = 0$ is rejected for large values of the test statistic

$$\frac{n - \text{rank}(X)}{\text{rank}(K)} \; \frac{\left(\hat{\boldsymbol{\theta}}_{(K)}\right)^T \left(K^T (X^T X)^- K\right)^- \hat{\boldsymbol{\theta}}_{(K)}}{\boldsymbol{Y}^T \left(I_n - X (X^T X)^- X^T\right) \boldsymbol{Y}}$$

where $\hat{\boldsymbol{\theta}}_{(K)} = K^T (X^T X)^- X^T \boldsymbol{Y}$.

With the gaussian assumption, under the null hypothesis the sampling distribution of the test statistic is a noncentral F with $(\text{rank}(K), n - \text{rank}(X))$ degrees of freedom and noncentrality parameter

$$\frac{1}{\sigma^2} \, (k^T\boldsymbol{\theta})^T \left(K^T (X^T X)^- K\right)^- (K^T\boldsymbol{\theta}) \, .$$

It is readily established that the power function of the F-test for the hypothesis $H_0 \colon K^T\boldsymbol{\theta} = 0$ is an increasing function of the noncentrality parameter.

2. CANONICAL MOMENTS

Under the assumption of gaussian "errors" $\varepsilon \frown \mathit{Gaussian}(0, \sigma^2)$, or even of a less demanding *moments assumption* involving homocedasticity, the choice of the allocation of the observations to accomplish the estimation in an optimal way amounts to dealing with the minimization of some functionals of the covariance matrix, and an elegant solution is provided using the theory of canonical moments and of closely related parameters (Dette and Studden, 1997):

Let

$$m_k(\mu) \;=\; m_k := \int_a^b x^k \, \mathrm{d}\mu(x) \;, \qquad k = 1, 2, ...$$

denote the k-th raw moment of the probability measure μ defined on the Borel sets of $[a, b]$, let

$$\boldsymbol{m}_n(\mu) \;=\; \boldsymbol{m}_n := (m_1, ..., m_n)$$

denote the vector of raw moments up to order n, and $\mathcal{P}_{\boldsymbol{m}}$ the class of all probability measures defined on the Borel sets of $[a, b]$ whose moments up to the order n are $m_1, ..., m_n$.

Skibinski (1967) investigated $m_{n+1}^{+} := \max\limits_{\mu \in \mathcal{P}_{\boldsymbol{m}}} \{m_{n+1}(\mu)\}$ and $m_{n+1}^{-} := \min\limits_{\mu \in \mathcal{P}_{\boldsymbol{m}}} \{m_{n+1}(\mu)\}$; from those "extreme" moments we can define several parameters, namely the *canonical moments*

$$\chi_k := \frac{m_k - m_{n+1}^{-}}{m_{n+1}^{+} - m_{n+1}^{-}} \;, \qquad k = 1, 2, ... \;,$$

and the closely associated parameters

$$\zeta_0 := 1 \;, \qquad \zeta_1 := \chi_1 \;, \qquad \zeta_k := \xi_{k-1} \, \chi_k \;, \quad k \geq 2 \;,$$

and

$$\gamma_0 := 1 \;, \qquad \gamma_1 := \eta_1 \;, \qquad \gamma_k := \chi_{k-1} \, \xi_k \;, \quad k \geq 2 \;,$$

where $\xi_k := 1 - \chi_k$; they have the substantial advantage of being invariant under linear transformations of the measure μ. From this invariance property, we shall in general consider $[a, b] = [-1, 1]$, or, whenever more appropriate, $[a, b] = [0, 1]$. Dette and Studden (1997, p. 21) claim that the parameters ζ_k and γ_k are more basic than the canonical moments.

The above parameters can be easily expressed in terms of the *Hankel determinants*

$$\underline{H}_{2n} := \begin{vmatrix} m_0 & \cdots & m_n \\ \vdots & \ddots & \vdots \\ m_n & \cdots & m_{2n} \end{vmatrix} \qquad \overline{H}_{2n} := \begin{vmatrix} m_1 - m_2 & \cdots & m_n - m_{n+1} \\ \vdots & \ddots & \vdots \\ m_n - m_{n+1} & \cdots & m_{2n-1} - m_{2n} \end{vmatrix}$$

and

$$\underline{H}_{2n+1} := \begin{vmatrix} m_1 & \cdots & m_{n+1} \\ \vdots & \ddots & \vdots \\ m_{n+1} & \cdots & m_{2n+1} \end{vmatrix} \qquad \overline{H}_{2n+1} := \begin{vmatrix} m_0 - m_1 & \cdots & m_n - m_{n+1} \\ \vdots & \ddots & \vdots \\ m_n - m_{n+1} & \cdots & m_{2n} - m_{2n+1} \end{vmatrix}$$

provided we define $\underline{H}_{-2} = \overline{H}_{-2} = \underline{H}_{-1} = \overline{H}_{-1} = \underline{H}_0 = \overline{H}_0 := 1$:

$$\chi_n = \frac{\underline{H}_n\, \overline{H}_{n-2}}{\underline{H}_{n-1}\, \overline{H}_{n-1}}, \quad \xi_n = \frac{\underline{H}_{n-2}\, \overline{H}_n}{\underline{H}_{n-1}\, \overline{H}_{n-1}}, \quad \zeta_n = \frac{\underline{H}_n\, \underline{H}_{n-3}}{\underline{H}_{n-1}\, \underline{H}_{n-2}}, \quad \gamma_n = \frac{\overline{H}_n\, \overline{H}_{n-3}}{\overline{H}_{n-1}\, \overline{H}_{n-2}}\ .$$

For instance, the canonical moments of $X \frown Beta(p,q)$, $p,q>0$, are $\chi_n = \left(\frac{1-(-1)^n}{2}\, p + \left[\frac{n}{2}\right]\right) / (p+q+n-1)$, $n=1,2,...$ (as usual, $[x]$ is the greatest integer less than or equal to x); observe, in particular, thal all the canonical moments of the $Beta\left(\frac{1}{2},\frac{1}{2}\right)$ (or arcsine) measure are $\chi_n = \frac{1}{2}$ (Skibinski, 1969).

It can be readily established that:

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & 0 & 1 \\ \dfrac{\chi_2}{2} & \xi_2 & \dfrac{\chi_2}{2} \end{cases}\ .$$

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, \chi_4, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & -\sqrt{\chi_2\,\xi_4} & \sqrt{\chi_2\,\xi_4} & 1 \\ \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{1}{2} - \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{1}{2} - \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} \end{cases}\ .$$

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, \chi_4, \frac{1}{2}, \chi_6, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & -\sqrt{\chi_2\,\xi_4 + \chi_4\,\xi_6} & 0 & \sqrt{\chi_2\,\xi_4 + \chi_4\,\xi_6} & 1 \\ \alpha_1 & \alpha_2 & 1 - 2\alpha_1 - 2\alpha_2 & \alpha_2 & \alpha_1 \end{cases},$$

where $\alpha_1 = \dfrac{\chi_2\,\chi_4\,\chi_6}{2(\xi_2\,\xi_4 + \chi_4\,\chi_6)}$ and $\alpha_2 = \dfrac{\chi_2\,\xi_4\,\xi_6}{2(\chi_2\,\xi_4 + \chi_4\,\xi_6)\,(\xi_2\,\xi_4 + \chi_4\,\chi_6)}$.

For a thorough discussion on moment spaces, moment sequences, canonical moments and their connection with Stieltjes transforms, continued fractions and orthogonal polynomials, cf. Dette and Studden (1997).

3. EXACT, APPROXIMATE AND OPTIMAL DESIGNS

In what follows, we shall assume that the unknown regression functions are sufficiently smooth over the range under investigation, so that modeling with a low degree polynomial $P_r(x) = \sum_{k=0}^{r} \theta_k x^k$ is appropriate.

In other words, $\boldsymbol{f}(x) = (1, x, ..., x^r)^T$, $k = r+1$, and if the observations are taken at the points $x_1, ..., x_n$, the design matrix is

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^r \\ 1 & x_2 & \cdots & x_2^r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^r \end{bmatrix} .$$

The design matrix X has rank $r+1$ if and only if there are at least $r+1$ different points among $x_1, ..., x_n$. We define the matrix of empirical moments up to order $2r$:

$$\frac{1}{n} X^T X = \begin{bmatrix} 1 & m_1 & m_2 & \cdots & m_r \\ m_1 & m_2 & m_3 & \cdots & m_{r+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_r & m_{r+1} & m_{r+2} & \cdots & m_{2r} \end{bmatrix}$$

with $m_i = \frac{1}{n} \sum_{k=0}^{n} x_k^i$, $i = 0, ..., 2r$.

The covariance matrix of the BLUE for the parameter subsystem $K^T\boldsymbol{\theta}$, where $K \in \mathbb{R}^{(r+1)\times s}$, is given by $\sigma^2 K^T (X^T X)^- K$.

If the experimenter is interested in finding out whether a polynomial regression of degree r or $r-1$ is appropriate for describing the response variable in terms of the explanatory variable, he can perform a F test as described above:

$$H_0 : \; K^T\boldsymbol{\theta} = \theta_r = 0$$

where $K = e_r = (0, 0, ..., 1)^T \in \mathbb{R}^{r+1}$ denotes the $(r+1)$-th unit vector. Assuming that the range inclusion is verified, in other words that there are at least $r+1$ different points among the $x_1, ..., x_n$, the test statistic

$$\frac{(n-r-1)\; \hat{\theta}^2_{r(e_r)} \left(e_r^T (X^T X)^{-1} e_r\right)^{-1}}{\boldsymbol{Y}^T \left(I_n - X (X^T X)^{-1} X^T\right) \boldsymbol{Y}} ,$$

where $\hat{\theta}_{r(e_r)} = e_r^T \hat{\boldsymbol{\theta}}^{GM}$, has under the null hypothesis the F distribution with $(1, n-m-1)$ degrees of freedom and noncentrality parameter $\frac{1}{\sigma^2} \theta_r^2 \left(e_r^T (X^T X)^{-1} e_r\right)^{-1}$.

As we observed above, the power function of the F-test for the null hypothesis H_0: $\theta_r = 0$ increases when $e_r^T (X^T X)^{-1} e_r$ decreases with respect to the choice of observation points — and this clearly raises the question whether there exists an optimum experimental design.

To discuss this issue, let us consider the linear regression model with the moment assumptions $\mathbb{E}[\boldsymbol{Y}] = X\boldsymbol{\theta}$ and $D(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_n$, where the design matrix is $X = \big(\boldsymbol{f}(x_1), ..., \boldsymbol{f}(x_n)\big)^T \in \mathbb{R}^{n \times k}$.

An *exact design* for sample size n is a finite probability measure on the design space $\mathcal{X}$ with support in the distinct points $x_1, ..., x_\ell$ among the $x_1, ..., x_n$, $\ell \leq n$, with masses $\dfrac{n_i}{n}$, $i = 1, ..., \ell$, that are multiples of $\dfrac{1}{n}$; n_i, $i = 1, ..., \ell$, is the number of times the particular point x_i occurs among $x_1, ..., x_n$. An exact design $\partial_{(n)}$ can therefore be represented

$$\partial_{(n)} = \begin{Bmatrix} x_1 & \cdots & x_\ell \\ \dfrac{n_1}{n} & \cdots & \dfrac{n_\ell}{n} \end{Bmatrix}$$

(Kiefer, 1959), and the matrix $X^T X$ is

$$\begin{aligned} X^T X \;&=\; \sum_{k=1}^{n} \boldsymbol{f}(x_k)\, \boldsymbol{f}^T(x_k) \;=\; n \sum_{j=1}^{\ell} \frac{n_i}{n}\, \boldsymbol{f}(x_j)\, \boldsymbol{f}^T(x_j) \\ &=\; n \int \boldsymbol{f}(x)\, \boldsymbol{f}^T(x)\, \mathrm{d}\partial_{(n)}(x) \;=:\; n\, \mathcal{M}(\partial_{(n)}) \;. \end{aligned}$$

Let $K \in \mathbb{R}^{k \times s}$ be a given matrix of rank $s \leq k$, and consider the problem of estimating the estimable parameter subsystem $K^T \boldsymbol{\theta}$; as the minimum dispersion matrix $D(\hat{\boldsymbol{\theta}}_{(K)}) = \frac{\sigma^2}{n} K^T M^{-1}(\partial_{(n)}) K$ depends on the design $\partial_{(n)}$, it is reasonable to choose an optimum exact design, whenever feasible, i.e. an exact design that for some optimality criterion minimizes the dispersion matrix.

Integer optimization raises many problems, and an approximate solution can be satisfactory. Hence it may be much more convenient to use an *approximate design*, defined as a probability measure on the design space $\mathcal{X}$ with support points $x_1, ..., x_\ell$ and weights $w_1, ..., w_\ell$ adding up to 1:

$$\partial_{(n)} = \begin{Bmatrix} x_1 & \cdots & x_\ell \\ w_1 & \cdots & w_\ell \end{Bmatrix} .$$

The interpretation is obvious, and exact designs for finite sample sizes can be found by apportionment from the optimal approximate designs (Fedorov, 1972), with the huge advantage that we can use the duality theory of convex analysis in the optimization of a concave function on a convex and compact subset of the set of nonnegative definite $s \times s$ matrices $NND(s)$ instead of integer optimization.

Pukelsheim (1993) discusses in depth several different optimality criteria or information functions — real valued, positively homogeneous, nonconstant, upper semicontinuous, isotonic and concave functions on $NND(s)$ — for determining optimum designs maximizing appropriate functions of the information matrix

$$C_{K}\big(M(\partial_{(n)})\big) =: \big(K^{T} M^{-1}(\partial_{(n)})\, K\big)^{-1}.$$

A design ∂^{*} is G-optimal for the parameter $\boldsymbol{\theta}$ if $|\mathcal{M}(\partial^{*})| > 0$ and it minimizes $G(\partial) = \max\limits_{x\in\mathcal{X}} \boldsymbol{f}^{T}(x)\, \mathcal{M}^{-1}(\partial)\, \boldsymbol{f}(x)$. G-optimal designs for low order polynomials have been first worked out numerically by Smith (1918), and theoretically by Guest (1958).

Hoel (1958) introduced D-optimal designs, the case $p = 0$ of Kiefer's ϕ_{p}-criteria we shall focus on, based on the definition of the p-th matrix mean

$$\phi_{p}(C) = \begin{cases} \lambda_{\min}(C) & p = -\infty \\ \big(\det\, C\big)^{\frac{1}{s}} & p = 0 \\ \left(\dfrac{1}{s}\ \mathrm{trace}\ C^{p}\right)^{\frac{1}{p}} & p \in (-\infty, 0) \cup (0, 1) \end{cases}$$

for $C \in PD(s)$, the set of positive definite $s \times s$ matrices, and

$$\phi_{p}(C) = \begin{cases} 0 & p \in [-\infty, 0] \\ \left(\dfrac{1}{s}\ \mathrm{trace}\ C^{p}\right)^{\frac{1}{p}} & p \in (0, 1] \end{cases}$$

for $C \in NND(s)$.

The popular D-optimality criterion uses $p = 0$:

$$\phi_{0}\Big(C_{K}\big(M(\partial_{(n)})\big)\Big) = \Big(\det\big(K^{T} M^{-1}(\partial_{(n)})\, K\big)\Big)^{-\frac{1}{s}}.$$

A D-optimum design $\partial^{D}_{(n)}$ for $K^{T}\boldsymbol{\theta}$ minimizes the volume of the ellipsoids of concentration for the vector $K^{T}\boldsymbol{\theta}$ with respect to the choice of designs $\partial_{(n)}$. In particular, if $K = I_{k}$, the D-optimum design $\partial^{D}_{(n)}$ maximizes $\det(\mathcal{M}(\partial_{(n)}))$.

Guest (1958) G-optimal designs and Hoel (1958) D-optimal designs coincide, and in 1960 Kiefer and Wolfowitz established the earliest "equivalence theorem": A design ∂^{*} with $\mathcal{M}(\partial^{*}) > 0$ is G-optimal for the parameter $\boldsymbol{\theta}$ if and only if it is D-optimal.

In what concerns the univariate polynomial regression model, Guest (1958) and Hoel (1958) results can be rephrased by noting that

$$|\mathcal{M}_{r}(\partial)| \;=\; \left|\int_{0}^{1} \boldsymbol{f}_{r}(x)\, \boldsymbol{f}_{r}^{T}(x)\ \mathrm{d}\partial(x)\right| \;=\; \underline{H}_{2r} = \prod_{j=1}^{r} (\zeta_{2j-1}\, \zeta_{2j})^{r-j+1}$$

and therefore

The D-optimal design ∂_r^D for the full parameter $\boldsymbol{\theta}$ in the univariate polynomial regression model of degree r on the interval $[-1,1]$ has equal masses at the $r+1$ zeros of the polynomial $(x^2-1)\,L_r'(x)$, where L_r' denotes the derivative of the r-th Legendre polynomial.

A D-optimal design on the interval $[a,b]$ is obviously obtained from ∂_r^D by the linear transformation $\partial_{[a,b]}(\{x\}) = \partial\left(\left\{\dfrac{2x-b-a}{b-a}\right\}\right)$. Observe also that $(x^2-1)\,L_r'(x) = r\,x\,L_r(x) - r\,L_{r-1}(x)$. Hence, for low degree polynomials, the optimal observation points are:

r									
2	-1	0	1						
3	-1	-0.44721	0.44721	1					
4	-1	-0.65465	0	0.65465	1				
5	-1	-0.76506	-0.28523	0.28523	0.76506	1			
6	-1	-0.83022	-0.46885	0	0.46885	0.83022	1		
7	-1	-0.8717	-0.59170	-0.20930	0.20930	0.59170	0.8717	1	
8	-1	-0.8998	-0.67719	-0.36312	0	0.36312	0.67719	0.8998	1

The D-efficiency of a given design in the polynomial regression of degree r is

$$\mathrm{eff}_r^D(\partial) = \left(\frac{|\mathcal{M}_r(\partial)|}{|\mathcal{M}_r(\partial^D)|}\right)^{\frac{1}{r+1}}.$$

On the other hand, the information for the parameter $K^T\boldsymbol{\theta} = \theta_r$ is given by

$$C_{e_r}\big(\mathcal{M}(\partial)\big) = \big(e_r^T\,\mathcal{M}_r^{-1}(\partial)\,e_r\big)^{-1} = \frac{|\mathcal{M}_r(\partial)|}{|\mathcal{M}_{r-1}(\partial)|}\,.$$

A design maximizing $C_{e_r}\big(\mathcal{M}(\partial)\big)$ is called D_1-optimal in the sense that it is optimal for the estimation of the highest coefficient θ_r:

The D_1-optimal design $\partial_r^{D_1}$ in the univariate polynomial regression of degree r on the interval $[-1,1]$ has equal masses $\dfrac{1}{2r}$ at the points -1 and 1, and equal masses $\dfrac{1}{r}$ at the zeros of the Chebyshev polynomial of second kind $U_{r-1}(x)$.

An example: In order to investigate if the quadratic term is relevant in the univariate quadratic model $\mathbf{Y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \boldsymbol{\varepsilon}$ on the design space $\mathcal{X} = [-1,1]$, we consider $K = \boldsymbol{e}_2 = (0,0,1)^T$.

Denoting $\partial_{(n)}$ an exact design of sample size n, and $\boldsymbol{f}(x) = (1, x, x^2)^T$ the vector of regression functions, the matrix $\boldsymbol{M}(\partial_{(n)})$ is

$$\boldsymbol{M}(\partial_{(n)}) = \int_{-1}^{1} \boldsymbol{f}(x)\, \boldsymbol{f}(x)^T \, \mathrm{d}\partial_{(n)}(x) = \begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix} .$$

The parameter $\theta_2 = \boldsymbol{e}_2^T \boldsymbol{\theta}$ is estimable if and only if $\xi_{(n)}$ has at least three support points, and for these designs the dispersion of the Gauss–Markov estimator is proportional to

$$\Big\{ C_K\big(\boldsymbol{M}(\partial_{(n)})\big) \Big\}^{-1} = \boldsymbol{e}_2^T \Big\{ \boldsymbol{M}(\partial_{(n)}) \Big\}^{-1} \boldsymbol{e}_2 = \frac{m_2 - m_1^2}{\big|\boldsymbol{M}(\partial_{(n)})\big|} .$$

The optimal designs, maximizing $C_K\big(\boldsymbol{M}(\partial_{(n)})\big)$ — and therefore minimizing the variance of the Gauss–Markov estimator of the parameter of interest θ_2 — in the set of all exact designs with nonsingular matrix $\boldsymbol{M}(\partial_{(n)})$ are

$$\partial^*_{(n)} = \begin{cases} \begin{cases} -1 & 0 & 1 \\ \dfrac{1}{4} & \dfrac{1}{2} & \dfrac{1}{4} \end{cases} & \text{if} \quad n = 4p \\ \begin{cases} -1 & 0 & 1 \\ \dfrac{p}{4p+1} & \dfrac{2p+1}{4p+1} & \dfrac{p}{4p+1} \end{cases} & \text{if} \quad n = 4p+1 \\ \begin{cases} -1 & \pm\, x_0(n) & 1 \\ \dfrac{p+1}{4p+2} & \dfrac{2p+1}{4p+2} & \dfrac{p}{4p+2} \end{cases} & \text{if} \quad n = 4p+2 \\ \begin{cases} -1 & 0 & 1 \\ \dfrac{p+1}{4p+3} & \dfrac{2p+1}{4p+3} & \dfrac{p+1}{4p+3} \end{cases} & \text{if} \quad n = 4p+3 \end{cases} ,$$

where in the case $n = 2p+2$ the point $x_0(n)$ is the real root of the cubic polynomial $n^2 x^3 - 3n x^2 + (n^2 - 2)\, x - n$ (Kraft and Schaefer, 1995).

On the other hand, an optimal approximate design to estimate θ_2 maximizes

$$C_{\boldsymbol{e}_2}\big(\mathcal{M}(\partial)\big) = \frac{|\mathcal{M}(\partial)|}{m_2 - m_1^2} = \frac{\underline{H}_4(\partial)}{\underline{H}_2(\partial)} .$$

This can be reexpressed in terms of the canonical moments of the measure ∂:

$$C_{\boldsymbol{e}_2}\big(\mathcal{M}(\partial)\big) = 2^4 \prod_{k=1}^{2} \gamma_{2k} = 2^4\, \chi_4 \prod_{j=1}^{3} \chi_j\, \xi_j .$$

The maximization in terms of canonical moments yelds $\chi_1 = \chi_2 = \chi_3 = \frac{1}{2}$ and $\chi_4 = 1$, and the approximate optimal design for estimating θ_2 is

$$\partial^* = \begin{cases} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{cases} .$$

Hence, n_0 denoting the closest integer to $\frac{n}{4}$, ∂^* approximates the exact design $\tilde{\partial}_{(n)} = \begin{cases} -1 & 0 & 1 \\ \frac{n_0}{n} & 1 - \frac{2n_0}{n} & \frac{n_0}{n} \end{cases}$.

In fact, they coincide unless $n = 4p + 2$, and in this case comparing the performance of the two designs using the relative efficiency ratio $\frac{C_{e_2}\left(\mathcal{M}(\tilde{\partial}_{(4p+2)})\right)}{C_{e_2}\left(\mathcal{M}(\partial^*_{(4p+2)})\right)}$ we can observe that for $p \geq 5$ we get $\frac{C_{e_2}\left(\mathcal{M}(\tilde{\partial}_{(4p+2)})\right)}{C_{e_2}\left(\mathcal{M}(\partial^*_{(4p+2)})\right)} \geq 0.995$, as seen on Table 1.

Table 1: Relative efficiency of the approximate design.

p	1	2	3	4	5
n	6	10	14	18	22
$x_0(p)$	0.0707	0.0408	0.0289	0.0224	0.0183
relative efficiency	0.9327	0.9759	0.9877	0.9925	0.9950

4. DISCRIMINANT, ROBUST AND MIXED DESIGNS

Consider the model $Y = \sum_{k=0}^{r} \theta_{rk} x^k + \varepsilon$, under the gaussian assumption. The optimal design to fit a linear regression model is fatally inefficient to detect curvature, and in general an optimal design for a specific task can be inappropriate for slightly different purposes. Hence we recommend that the analysis be performed in two steps, first to try to identify the appropriate degree of the polynomial, then to build up the optimal design.

The two steps can however be merged if practical considerations on data gathering costs imply that should be so.

Anderson (1962) invented a good decision rule for this problem: For a given nondecreasing sequence of levels $(\alpha_1, ..., \alpha_r)$ the procedure he devised chooses the largest integer in $\{1, ..., r\}$ for which the F-test rejects the null hypothesis H_0: $\theta_{jj} = 0$ at the levels α_j. This method has several optimality properties, and led to the introduction of discriminant and of robust designs, discussed in what follows.

Let $\mathcal{F}_r$ be the class of all possible polynomial regression models up to degree r, and $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$ nonnegative numbers with $\pi_r > 0$ and such that $\pi_1 + \cdots + \pi_r = 1$. Those are interpreted as "priors" reflecting the experimenter belief about the adequacy of the polynomial regression of degree ℓ, $\ell = 1, ..., r$.

As discussed beforehand, H_0: $\theta_{\ell\ell} = 0$ can be tested using a test statistic with non-central F distribution, and the power function increases with the non centrality parameter which we now rewrite $\delta_\ell^2(\partial) = \frac{\theta_{\ell\ell}^2}{\sigma^2} \left(e_\ell^T \left(X_l^T X_\ell\right)^{-1} e_\ell\right)^{-1}$. As this should ideally be maximized for $\ell = 1, ..., r$, which would amount to jointly maximizing

$$\text{eff}_\ell^{D_1}(\partial) \;=\; \frac{\delta_\ell^2(\partial)}{\sup\limits_\eta \delta_\ell^2(\eta)} \;=\; 2^{2\ell-2}\, \frac{|M_\ell(\partial)|}{|M_{\ell-1}(\partial)|}\;,$$

a task obviously beyond what is feasible, what can be done in practice is to maximize an appropriate weighted mean of the above efficiencies, using the weights in $\boldsymbol{\pi}$ corresponding to the credibility the experimenter puts in the adequacy of using polynomal regression of each of the degrees ℓ, $\ell = 1, ..., r$.

*A design $\partial_{0,\boldsymbol{\pi}}$ with moment matrix $\mathcal{M}(\partial_{0,\boldsymbol{\pi}})$ is a Ψ_0-**optimal discriminating design** for the class $\mathcal{F}_r$ with respect to the prior $\boldsymbol{\pi}$ if and only if $\partial_{0,\boldsymbol{\pi}}$ maximizes the weighted geometric mean*

$$\Psi_0^{\boldsymbol{\pi}}(\partial) \;=\; \prod_{k=1}^{r} \left(\text{eff}_k^{D_1}(\partial)\right)^{\boldsymbol{\pi}_k} = \prod_{k=1}^{r} \left(\frac{2^{4k-2}}{(b-a)^{2k}}\; \frac{|\mathcal{M}_k(\partial)|}{|\mathcal{M}_{k-1}(\partial)|} \right)^{\boldsymbol{\pi}_k}.$$

(Observe that if $\boldsymbol{\pi} = (0, ..., 0, 1)$ we obtain the D_1 optimality criterion.)

It is readily established that the Ψ_0-optimal discriminating design for the class $\mathcal{F}_r$ with respect to the prior $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$ is uniquely determined by its canonical moments

$$\chi_{2i-1} = \frac{1}{2}, \quad i = 1, ..., r\;, \qquad \chi_{2i} = \frac{\Pi_i}{\Pi_i + \Pi_{i+1}}, \quad i = 1, ..., r-1\;, \qquad \chi_{2r} = 1\;,$$

where $\Pi_i = \sum\limits_{\ell=i}^{r} \pi_\ell$ (Lau and Studden, 1985). For instance, with the uniform prior $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ for the class $\mathcal{F}_4$ we have

$$\Pi_1 = 1\,, \quad \Pi_2 = \frac{3}{4}, \quad \Pi_3 = \frac{1}{2}, \quad \Pi_4 = \frac{4}{4}\;,$$

and

$$\chi_2 = \frac{4}{7}, \quad \chi_4 = \frac{3}{5}, \quad \chi_6 = \frac{2}{3}\;.$$

Therefore the the Ψ_0-optimal discriminating design is

$$\partial_{0,\boldsymbol{\pi}_U} = \begin{cases} -1 & -\sqrt{\frac{3}{7}} & 0 & \sqrt{\frac{3}{7}} & 1 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{cases}.$$

In what concerns $\Psi_0^{\boldsymbol{\pi}}$-optimal discriminant designs for the classes $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$, and with $\boldsymbol{\pi}$ giving the same prior probability $1/r$ to the values of ℓ ranging from 1 to r,

r	$\boldsymbol{\pi}$	Points	$\text{eff}_1^{D_1}$/ $\text{eff}_2^{D_1}$/ $\text{eff}_3^{D_1}$/ $\text{eff}_4^{D_1}$ (ξ)
2	$\left(\frac{1}{2}, \frac{1}{2}\right)$	-1, 0, 1	0.817 / 1 / $-$ / $-$
3	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	-1, -0.4472, 0, 4472, 1	0.600 / 0.640 / 0.853 / $-$
4	$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$	-1, -0.6547, 0, 0.6547, 1	0.571 / 0.588 / 0.627 / 0.836

we can observe, when comparing with the efficiency of the D_1-optimal design for polynomial regression of degree $r=4$, that the loss of efficiency in the case of degree 4 is largely compensated by the increased efficiency when the appropriate degree is lower than 4.

An alternative strategy, inspired on the way $\Psi_0^{\boldsymbol{\pi}}$-optimal discriminant designs have been defined, is to build up designs maximizing an weighted geometric mean of D-efficiencies, up to some degree r. Those designs are christened robust designs since they are quite efficient for a set of possible polynomial regression degrees.

For a given weights vector $\boldsymbol{\pi}=(\pi_1,...,\pi_r)$, $\sum_{l=1}^r \pi_l = 1$ *and* $\pi_i > 0$, *the design* $\partial_{0,\boldsymbol{\pi}}$ *is a* $\Xi_0^{\boldsymbol{\pi}}$***-robust design*** *for the class* $\mathcal{F}_r$ *in respect to the prior* $\boldsymbol{\pi}$ *if and only if* $\partial_{0,\boldsymbol{\pi}}$ *maximizes the weighted geometric mean*

$$\Xi_0^{\boldsymbol{\pi}}(\partial) = \prod_{\ell=1}^{r} \left(\text{eff}_\ell^D(\partial)\right)^{\pi_\ell} = \prod_{\ell=1}^{r} \left(\frac{|M_\ell(\partial)|}{|M_\ell(\partial^D)|}\right)^{\pi_\ell/(\ell+1)} .$$

Dette and Studden (1995) show that the canonical moments for the above defined robust design are

$$\chi_{2i-1} = \frac{1}{2}, \quad i=1,...,r\ , \qquad \chi_{2i} = \frac{\sigma_i}{\sigma_i+\sigma_{i+1}}, \quad i=1,...,r-1\ , \qquad \chi_{2r}=1\ ,$$

with $\sigma_i = \sum_{\ell=i}^{r} \frac{\ell+1-i}{\ell+1}\,\pi_\ell$.

For $\Xi_0^{\boldsymbol{\pi}}$-robust designs for the classes $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$, and with $\boldsymbol{\pi}$ giving the same prior probability $1/r$ to the values of ℓ ranging from 1 to r,

m	Points	Weights	$\text{eff}_1^D/\text{eff}_2^D/\text{eff}_3^D/\text{eff}_4^D$ (ξ)
2	$-1, 0, 1$	$0.389, 0.222, 0, 389$	$0.881/0.968/-/-$
3	$-1, -0.401, 0.401, 1$	$0.319, 0.181, 0.181, 0.319$	$0.835/0.914/0.954/-$
4	$-1, -0.605, 0, -1, 0.605, 1$	$0.271, 0.152, 0.153, 0.152, 0.271$	$0.809/0.883/0.927/0.949$

As we shall show in Tables 5–12 below, gross loss of efficiency can be incurred into — up to 10% — when a $\Xi_0^{\boldsymbol{\pi}}$-robust design is used instead of a $\Psi_0^{\boldsymbol{\pi}}$-discriminant design, or vice-versa. This prompted us to use a mixed strategy, defining $\Theta_0^{\boldsymbol{\pi}}$-mixed designs as follows:

For a given weights vector $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$, $\sum_{l=1}^r \pi_l = 1$ *and* $\pi_i > 0$, *the design* $\partial_{0,\boldsymbol{\pi}}$ *is a* $\Theta_0^{\boldsymbol{\pi}}$***-mixed design*** *for the class* $\mathcal{F}_r$ *in respect to the prior* $\boldsymbol{\pi}$ *if and only if* $\partial_{0,\boldsymbol{\pi}}$ *maximizes the weighted geometric mean*

$$\begin{aligned}\Theta_0^{\boldsymbol{\pi}} &= \prod_{\ell=1}^r \big(\mathrm{eff}_\ell^{D_1}(\partial)\big)^{\pi_\ell} \prod_{j=1}^r \big(\mathrm{eff}_j^{D}(\partial)\big)^{\pi_j} \\ &= \prod_{\ell=1}^r \left(2^{2\ell-2}\frac{|M_\ell(\partial)|}{|M_{\ell-1}(\partial)|}\right)^{\pi_\ell} \prod_{j=1}^r \left(\frac{M_j(\partial)}{M_j(\partial_j^D)}\right)^{\frac{\pi_j}{j+1}}.\end{aligned}$$

In Tables 2–4 we present mixed designs for $\mathcal{F}_r$, $r = 2, 3, 4$, and in Tables 5–12 we study the corresponding efficiencies when they are used instead of the corresponding optimal discriminant or robust designs.

Table 2: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r = 2$, $\boldsymbol{\pi} = (a, 1-a)$.

a	weight at ± 1	weight at 0
0.05	0.2835	0.4330
0.10	0.2895	0.4211
0.15	0.2958	0.4084
0.20	0.3025	0.3951
0.25	0.3095	0.3810
0.30	0.3170	0.3660
0.35	0.3249	0.3502
0.40	0.3333	0.3333
0.45	0.3423	0.3154
0.50	0.3519	0.2963
0.55	0.3621	0.2759
0.60	0.3730	0.2540
0.65	0.3848	0.2305
0.70	0.3974	0.2051
0.75	0.4111	0.1778
0.80	0.4259	0.1481
0.85	0.4420	0.1159
0.90	0.4596	0.0808
0.95	0.4788	0.0423

Table 3: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=3$, $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	t	weight at ± 1	weight at $\pm t$
0.1	0.1	0.4911	0.2119	0.2881
0.1	0.2	0.4748	0.2190	0.2810
0.1	0.3	0.4553	0.2267	0.2733
0.1	0.4	0.4315	0.2350	0.2650
0.1	0.5	0.4019	0.2440	0.2560
0.1	0.6	0.3635	0.2538	0.2462
0.1	0.7	0.3112	0.2646	0.2354
0.1	0.8	0.2318	0.2764	0.2236
0.2	0.1	0.5001	0.2256	0.2744
0.2	0.2	0.4808	0.2338	0.2662
0.2	0.3	0.4569	0.2428	0.2572
0.2	0.4	0.4269	0.2525	0.2475
0.2	0.5	0.3876	0.2632	0.2368
0.2	0.6	0.3333	0.2750	0.2250
0.2	0.7	0.2496	0.2880	0.2120
0.3	0.1	0.5095	0.2415	0.2585
0.3	0.2	0.4858	0.2513	0.2487
0.3	0.3	0.4556	0.2619	0.2381
0.3	0.4	0.4155	0.2736	0.2264
0.3	0.5	0.3593	0.2865	0.2135
0.3	0.6	0.2709	0.3009	0.1991
0.4	0.1	0.5190	0.2606	0.2394
0.4	0.2	0.4889	0.2722	0.2278
0.4	0.3	0.4483	0.2851	0.2149
0.4	0.4	0.3902	0.2994	0.2006
0.4	0.5	0.2967	0.3154	0.1846
0.5	0.1	0.5281	0.2836	0.2164
0.5	0.2	0.4875	0.2978	0.2022
0.5	0.3	0.4279	0.3137	0.1863
0.5	0.4	0.3289	0.3316	0.1684
0.6	0.1	0.5351	0.3121	0.1879
0.6	0.2	0.4748	0.3299	0.1701
0.6	0.3	0.3705	0.3500	0.1500
0.3	0.4	0.5353	0.3481	0.1519
0.3	0.5	0.4267	0.3710	0.1290
0.3	0.6	0.5085	0.3953	0.1047
1/3	1/3	0.4407	0.2731	0.2269

Table 4: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=4$, $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	t	weight at ± 1	weight at $\pm t$	weight at 0
0.1	0.1	0.1	0.6973	0.1717	0.2177	0.2210
0.1	0.1	0.2	0.6836	0.1764	0.2176	0.2119
0.1	0.1	0.3	0.6673	0.1814	0.2184	0.2004
0.1	0.1	0.4	0.6474	0.1867	0.2206	0.1853
0.1	0.1	0.5	0.6228	0.1924	0.2252	0.1647
0.1	0.1	0.6	0.5913	0.1985	0.2341	0.1349
0.1	0.1	0.7	0.5495	0.2050	0.2513	0.0874
0.1	0.2	0.1	0.6937	0.1811	0.2036	0.2307
0.1	0.2	0.2	0.6765	0.1864	0.2036	0.2200
0.1	0.2	0.3	0.6553	0.1920	0.2051	0.2057
0.1	0.2	0.4	0.6284	0.1981	0.2091	0.1856
0.1	0.2	0.5	0.5933	0.2046	0.2178	0.1553
0.1	0.2	0.6	0.5453	0.2115	0.2363	0.1043
0.1	0.3	0.1	0.6886	0.1917	0.1873	0.2420
0.1	0.3	0.2	0.6661	0.1977	0.1879	0.2289
0.1	0.3	0.3	0.6370	0.2042	0.1908	0.2100
0.1	0.3	0.4	0.5979	0.2111	0.1988	0.1802
0.1	0.3	0.5	0.5423	0.2186	0.2182	0.1265
0.1	0.4	0.1	0.6811	0.2038	0.1686	0.2552
0.1	0.4	0.2	0.6499	0.2107	0.1701	0.2383
0.1	0.4	0.3	0.6065	0.2182	0.1766	0.2104
0.1	0.4	0.4	0.5417	0.2262	0.1959	0.1557
0.1	0.5	0.1	0.6692	0.2177	0.1469	0.2707
0.1	0.5	0.2	0.6217	0.2258	0.1509	0.2466
0.1	0.5	0.3	0.5462	0.2345	0.1683	0.1943
0.1	0.6	0.1	0.6481	0.2340	0.1216	0.2887
0.1	0.6	0.2	0.5610	0.2435	0.1342	0.2447
0.1	0.7	0.1	0.5981	0.2533	0.0935	0.3064
0.2	0.1	0.1	0.7031	0.1851	0.2128	0.2041
0.2	0.1	0.2	0.6869	0.1907	0.2124	0.1937
0.2	0.1	0.3	0.6670	0.1967	0.2132	0.1801
0.2	0.1	0.4	0.6419	0.2032	0.2162	0.1613
0.2	0.1	0.5	0.6092	0.2101	0.2231	0.1335
0.2	0.1	0.6	0.5647	0.2175	0.2384	0.0881
0.2	0.2	0.1	0.6978	0.1964	0.1965	0.2143
0.2	0.2	0.2	0.6766	0.2028	0.1964	0.2017
0.2	0.2	0.3	0.6492	0.2097	0.1985	0.1837
0.2	0.2	0.4	0.6125	0.2171	0.2049	0.1559
0.2	0.2	0.5	0.5607	0.2251	0.2212	0.1074
0.2	0.3	0.1	0.6900	0.2093	0.1775	0.2265
0.2	0.3	0.2	0.6604	0.2167	0.1783	0.2101
0.2	0.3	0.3	0.6195	0.2247	0.1835	0.1836

(continued on next page)

Table 4: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=4$, $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	t	weight at ± 1	weight at $\pm t$	weight at 0
0.2	0.3	0.4	0.5586	0.2333	0.2001	0.1332
0.2	0.4	0.1	0.6778	0.2242	0.1552	0.2411
0.2	0.4	0.2	0.6325	0.2329	0.1584	0.2175
0.2	0.4	0.3	0.5610	0.2423	0.1737	0.1681
0.2	0.5	0.1	0.6562	0.2418	0.1291	0.2582
0.2	0.5	0.2	0.5728	0.2520	0.1404	0.2151
0.2	0.6	0.1	0.6058	0.2627	0.0997	0.2753
0.3	0.1	0.1	0.7089	0.2014	0.2058	0.1857
0.3	0.1	0.2	0.6891	0.2082	0.2049	0.1737
0.3	0.1	0.3	0.6637	0.2156	0.2060	0.1569
0.3	0.1	0.4	0.6297	0.2236	0.2107	0.1316
0.3	0.1	0.5	0.5820	0.2322	0.2234	0.0888
0.3	0.2	0.1	0.7009	0.2152	0.1865	0.1966
0.3	0.2	0.2	0.6732	0.2231	0.1864	0.1809
0.3	0.2	0.3	0.6349	0.2317	0.1902	0.1562
0.3	0.2	0.4	0.5785	0.2411	0.2035	0.1109
0.3	0.3	0.1	0.6884	0.2313	0.1639	0.2098
0.3	0.3	0.2	0.6457	0.2406	0.1658	0.1872
0.3	0.3	0.3	0.5787	0.2508	0.1784	0.1416
0.3	0.4	0.1	0.6663	0.2503	0.1369	0.2256
0.3	0.4	0.2	0.5872	0.2614	0.1464	0.1844
0.3	0.5	0.1	0.6155	0.2731	0.1061	0.2417
0.4	0.1	0.1	0.7144	0.2216	0.1957	0.1656
0.4	0.1	0.2	0.6888	0.2301	0.1944	0.1510
0.4	0.1	0.3	0.6536	0.2394	0.1963	0.1287
0.4	0.1	0.4	0.6022	0.2495	0.2059	0.0892
0.4	0.2	0.1	0.7016	0.2389	0.1726	0.1771
0.4	0.2	0.2	0.6619	0.2490	0.1730	0.1561
0.4	0.2	0.3	0.6001	0.2601	0.1823	0.1152
0.4	0.3	0.1	0.6791	0.2595	0.1449	0.1912
0.4	0.3	0.2	0.6051	0.2717	0.1520	0.1527
0.4	0.4	0.1	0.6281	0.2845	0.1126	0.2057
0.5	0.1	0.1	0.7182	0.2472	0.1811	0.1434
0.5	0.1	0.2	0.6821	0.2582	0.1796	0.1245
0.5	0.1	0.3	0.6263	0.2703	0.1850	0.0894
0.5	0.2	0.1	0.6956	0.2697	0.1527	0.1552
0.5	0.2	0.2	0.6276	0.2830	0.1566	0.1207
0.5	0.3	0.1	0.6445	0.2972	0.1190	0.1676
0.6	0.1	0.1	0.7172	0.2810	0.1598	0.1184
0.6	0.1	0.2	0.6566	0.2957	0.1597	0.0892
0.6	0.2	0.1	0.6667	0.3114	0.1246	0.1280
0.7	0.1	0.1	0.6978	0.3274	0.1285	0.0882
0.25	0.25	0.25	0.6484	0.2239	0.1839	0.1845

Table 5: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=2$ and $\boldsymbol{\pi}=(a,1-a)$.

a	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\big[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\big]$	$100\times$ $\big[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\big]$
0.05	0.967	0.866	0.955	−10.076	−1.110
0.10	0.935	0.844	0.926	−9.118	−0.995
0.15	0.907	0.824	0.898	−8.233	−0.890
0.20	0.880	0.806	0.872	−7.414	−0.795
0.25	0.856	0.790	0.849	−6.657	−0.707
0.30	0.834	0.775	0.828	−5.958	−0.627
0.35	0.814	0.761	0.809	−5.312	−0.554
0.40	0.797	0.750	0.792	−4.714	−0.487
0.45	0.782	0.741	0.778	−4.162	−0.426
0.50	0.770	0.733	0.766	−3.650	−0.370
0.55	0.760	0.728	0.757	−3.177	−0.320
0.60	0.753	0.726	0.751	−2.739	−0.273
0.65	0.750	0.727	0.748	−2.331	−0.230
0.70	0.751	0.731	0.749	−1.951	−0.191
0.75	0.757	0.741	0.755	−1.596	−0.155
0.80	0.768	0.756	0.767	−1.262	−0.121
0.85	0.789	0.779	0.788	−0.943	−0.090
0.90	0.822	0.815	0.821	−0.635	−0.060
0.95	0.877	0.873	0.876	−0.328	−0.031

Table 6: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=2$ and $\boldsymbol{\pi}=(a,1-a)$.

a	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\big[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\big]$	$100\times$ $\big[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\big]$
0.05	0.939	0.990	0.967	−5.084	−2.343
0.10	0.934	0.981	0.959	−4.679	−2.161
0.15	0.929	0.972	0.952	−4.293	−1.986
0.20	0.924	0.963	0.945	−3.923	−1.819
0.25	0.919	0.955	0.938	−3.570	−1.658
0.30	0.915	0.947	0.932	−3.233	−1.505
0.35	0.911	0.941	0.927	−2.912	−1.359
0.40	0.908	0.934	0.922	−2.607	−1.219
0.45	0.906	0.929	0.918	−2.317	−1.086
0.50	0.904	0.924	0.914	−2.042	−0.959
0.55	0.902	0.920	0.912	−1.781	−0.839
0.60	0.902	0.917	0.910	−1.534	−0.724
0.65	0.903	0.916	0.910	−1.301	−0.616
0.70	0.905	0.916	0.911	−1.081	−0.513
0.75	0.909	0.917	0.913	−0.873	−0.416
0.80	0.914	0.921	0.918	−0.678	−0.324
0.85	0.923	0.928	0.926	−0.495	−0.237
0.90	0.937	0.940	0.938	−0.321	−0.154
0.95	0.957	0.959	0.958	−0.158	−0.076

Table 7: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=3$ and $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$
0.1	0.1	0.875	0.772	0.866	−10.305	−0.976
0.1	0.2	0.825	0.736	0.816	−8.885	−0.862
0.1	0.3	0.783	0.705	0.775	−7.788	−0.778
0.1	0.4	0.751	0.681	0.744	−6.989	−0.721
0.1	0.5	0.730	0.665	0.723	−6.478	−0.688
0.1	0.6	0.722	0.659	0.715	−6.266	−0.683
0.1	0.7	0.733	0.669	0.726	−6.405	−0.708
0.1	0.8	0.776	0.706	0.769	−7.058	−0.781
0.2	0.1	0.824	0.741	0.817	−8.285	−0.758
0.2	0.2	0.777	0.707	0.771	−7.049	−0.665
0.2	0.3	0.740	0.679	0.734	−6.133	−0.600
0.2	0.4	0.714	0.659	0.708	−5.518	−0.561
0.2	0.5	0.700	0.648	0.695	−5.206	−0.548
0.2	0.6	0.704	0.652	0.699	−5.232	−0.564
0.2	0.7	0.739	0.681	0.733	−5.724	−0.621
0.3	0.1	0.781	0.716	0.776	−6.556	−0.580
0.3	0.2	0.738	0.684	0.733	−5.485	−0.504
0.3	0.3	0.706	0.659	0.702	−4.741	−0.456
0.3	0.4	0.686	0.643	0.682	−4.314	−0.435
0.3	0.5	0.683	0.641	0.679	−4.227	−0.442
0.3	0.6	0.709	0.663	0.704	−4.577	−0.487
0.4	0.1	0.747	0.696	0.743	−5.077	−0.435
0.4	0.2	0.708	0.666	0.704	−4.157	−0.373
0.4	0.3	0.681	0.645	0.678	−3.587	−0.342
0.4	0.4	0.671	0.637	0.667	−3.374	−0.340
0.4	0.5	0.687	0.651	0.683	−3.591	−0.373
0.5	0.1	0.722	0.684	0.719	−3.809	−0.316
0.5	0.2	0.687	0.657	0.685	−3.038	−0.269
0.5	0.3	0.668	0.642	0.666	−2.667	−0.255
0.5	0.4	0.675	0.647	0.672	−2.745	−0.278
0.6	0.1	0.708	0.681	0.706	−2.718	−0.219
0.6	0.2	0.679	0.657	0.677	−2.116	−0.188
0.6	0.3	0.674	0.654	0.672	−2.024	−0.198
0.7	0.1	0.708	0.690	0.706	−1.775	−0.141
0.7	0.2	0.689	0.675	0.688	−1.424	−0.131
0.8	0.1	0.729	0.719	0.728	−0.982	−0.080
1/3	1/3	0.689	0.648	0.685	−4.174	−0.407

Table 8: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=3$ and $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.1	0.1	0.913	0.962	0.938	−4.996	−2.433
0.1	0.2	0.905	0.951	0.929	−4.597	−2.208
0.1	0.3	0.898	0.941	0.920	−4.288	−2.029
0.1	0.4	0.892	0.932	0.914	−4.070	−1.898
0.1	0.5	0.887	0.927	0.909	−3.945	−1.817
0.1	0.6	0.885	0.924	0.907	−3.920	−1.790
0.1	0.7	0.887	0.927	0.909	−4.005	−1.823
0.1	0.8	0.897	0.939	0.920	−4.221	−1.929
0.2	0.1	0.900	0.941	0.921	−4.076	−2.005
0.2	0.2	0.893	0.930	0.912	−3.713	−1.797
0.2	0.3	0.887	0.922	0.905	−3.449	−1.641
0.2	0.4	0.882	0.915	0.900	−3.288	−1.538
0.2	0.5	0.880	0.912	0.897	−3.234	−1.492
0.2	0.6	0.881	0.914	0.898	−3.296	−1.510
0.2	0.7	0.889	0.924	0.908	−3.494	−1.603
0.3	0.1	0.890	0.922	0.906	−3.258	−1.620
0.3	0.2	0.884	0.913	0.899	−2.930	−1.429
0.3	0.3	0.879	0.906	0.893	−2.716	−1.297
0.3	0.4	0.875	0.902	0.889	−2.621	−1.227
0.3	0.5	0.876	0.902	0.890	−2.654	−1.226
0.3	0.6	0.882	0.911	0.898	−2.831	−1.305
0.4	0.1	0.881	0.906	0.894	−2.534	−1.273
0.4	0.2	0.876	0.899	0.888	−2.245	−1.101
0.4	0.3	0.873	0.894	0.884	−2.090	−0.999
0.4	0.4	0.872	0.893	0.883	−2.080	−0.973
0.4	0.5	0.878	0.900	0.890	−2.229	−1.034
0.5	0.1	0.875	0.894	0.884	−1.898	−0.963
0.5	0.2	0.872	0.888	0.880	−1.654	−0.814
0.5	0.3	0.871	0.887	0.879	−1.579	−0.752
0.5	0.4	0.875	0.892	0.884	−1.688	−0.789
0.6	0.1	0.873	0.887	0.880	−1.344	−0.687
0.6	0.2	0.872	0.884	0.878	−1.164	−0.571
0.6	0.3	0.876	0.888	0.882	−1.209	−0.572
0.7	0.1	0.877	0.886	0.882	−0.868	−0.446
0.7	0.2	0.881	0.889	0.885	−0.798	−0.385
0.8	0.1	0.892	0.897	0.895	−0.482	−0.244
1/3	1/3	0.875	0.900	0.888	−2.458	−1.166

Table 9: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$
0.1	0.1	0.1	0.820	0.722	0.811	−9.740	−0.814
0.1	0.1	0.2	0.773	0.688	0.766	−8.483	−0.741
0.1	0.1	0.3	0.736	0.660	0.729	−7.586	−0.693
0.1	0.1	0.4	0.710	0.640	0.703	−7.035	−0.668
0.1	0.1	0.5	0.696	0.628	0.690	−6.839	−0.668
0.1	0.1	0.6	0.700	0.630	0.693	−7.059	−0.696
0.1	0.1	0.7	0.735	0.656	0.727	−7.876	−0.768
0.1	0.2	0.1	0.772	0.691	0.765	−8.138	−0.698
0.1	0.2	0.2	0.730	0.659	0.724	−7.068	−0.639
0.1	0.2	0.3	0.698	0.634	0.692	−6.369	−0.607
0.1	0.2	0.4	0.679	0.618	0.673	−6.043	−0.599
0.1	0.2	0.5	0.676	0.614	0.669	−6.132	−0.619
0.1	0.2	0.6	0.701	0.633	0.694	−6.787	−0.681
0.1	0.3	0.1	0.734	0.665	0.728	−6.914	−0.620
0.1	0.3	0.2	0.696	0.636	0.690	−6.013	−0.574
0.1	0.3	0.3	0.669	0.614	0.664	−5.519	−0.555
0.1	0.3	0.4	0.659	0.604	0.653	−5.459	−0.566
0.1	0.3	0.5	0.675	0.616	0.669	−5.960	−0.617
0.1	0.4	0.1	0.705	0.644	0.699	−6.039	−0.575
0.1	0.4	0.2	0.671	0.618	0.665	−5.306	−0.540
0.1	0.4	0.3	0.652	0.602	0.647	−5.053	−0.538
0.1	0.4	0.4	0.659	0.605	0.653	−5.387	−0.577
0.1	0.5	0.1	0.686	0.631	0.681	−5.500	−0.559
0.1	0.5	0.2	0.658	0.608	0.652	−4.962	−0.537
0.1	0.5	0.3	0.653	0.602	0.647	−5.079	−0.561
0.1	0.6	0.1	0.681	0.627	0.675	−5.312	−0.574
0.1	0.6	0.2	0.662	0.612	0.657	−5.088	−0.573
0.1	0.7	0.1	0.695	0.640	0.689	−5.571	−0.625
0.2	0.1	0.1	0.772	0.694	0.766	−7.742	−0.622
0.2	0.1	0.2	0.730	0.663	0.724	−6.673	−0.565
0.2	0.1	0.3	0.698	0.638	0.692	−5.970	−0.532
0.2	0.1	0.4	0.678	0.622	0.673	−5.632	−0.524
0.2	0.1	0.5	0.675	0.618	0.670	−5.698	−0.541
0.2	0.1	0.6	0.700	0.637	0.694	−6.309	−0.596
0.2	0.2	0.1	0.729	0.665	0.723	−6.330	−0.525
0.2	0.2	0.2	0.691	0.636	0.686	−5.448	−0.483
0.2	0.2	0.3	0.664	0.615	0.660	−4.959	−0.466
0.2	0.2	0.4	0.654	0.605	0.649	−4.888	−0.477
0.2	0.2	0.5	0.670	0.617	0.665	−5.352	−0.523
0.2	0.3	0.1	0.695	0.642	0.690	−5.301	−0.467
0.2	0.3	0.2	0.661	0.615	0.657	−4.598	−0.437
0.2	0.3	0.3	0.643	0.599	0.638	−4.354	−0.437

(continued on next page)

Table 9: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\big[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\big]$	$100\times$ $\big[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\big]$
0.2	0.3	0.4	0.649	0.602	0.644	−4.663	−0.474
0.2	0.4	0.1	0.671	0.625	0.667	−4.632	−0.441
0.2	0.4	0.2	0.643	0.602	0.639	−4.131	−0.424
0.2	0.4	0.3	0.639	0.596	0.634	−4.244	−0.449
0.2	0.5	0.1	0.660	0.617	0.656	−4.330	−0.447
0.2	0.5	0.2	0.642	0.601	0.638	−4.135	−0.451
0.2	0.6	0.1	0.668	0.624	0.663	−4.468	−0.489
0.3	0.1	0.1	0.732	0.672	0.728	−6.028	−0.466
0.3	0.1	0.2	0.694	0.643	0.690	−5.135	−0.423
0.3	0.1	0.3	0.668	0.621	0.664	−4.630	−0.406
0.3	0.1	0.4	0.657	0.612	0.653	−4.536	−0.413
0.3	0.1	0.5	0.674	0.624	0.669	−4.961	−0.454
0.3	0.2	0.1	0.693	0.645	0.689	−4.790	−0.387
0.3	0.2	0.2	0.659	0.618	0.656	−4.094	−0.359
0.3	0.2	0.3	0.641	0.603	0.637	−3.844	−0.359
0.3	0.2	0.4	0.647	0.606	0.643	−4.124	−0.393
0.3	0.3	0.1	0.664	0.624	0.660	−3.952	−0.347
0.3	0.3	0.2	0.636	0.601	0.633	−3.471	−0.333
0.3	0.3	0.3	0.632	0.596	0.628	−3.573	−0.357
0.3	0.4	0.1	0.647	0.612	0.643	−3.511	−0.342
0.3	0.4	0.2	0.630	0.596	0.626	−3.336	−0.348
0.3	0.5	0.1	0.648	0.613	0.645	−3.523	−0.374
0.4	0.1	0.1	0.701	0.655	0.698	−4.557	−0.341
0.4	0.1	0.2	0.667	0.629	0.664	−3.838	−0.311
0.4	0.1	0.3	0.649	0.613	0.645	−3.560	−0.308
0.4	0.1	0.4	0.655	0.617	0.652	−3.802	−0.337
0.4	0.2	0.1	0.665	0.631	0.663	−3.484	−0.277
0.4	0.2	0.2	0.638	0.608	0.635	−3.000	−0.265
0.4	0.2	0.3	0.633	0.603	0.630	−3.080	−0.286
0.4	0.3	0.1	0.642	0.614	0.640	−2.859	−0.257
0.4	0.3	0.2	0.625	0.598	0.622	−2.690	−0.265
0.4	0.4	0.1	0.637	0.609	0.634	−2.724	−0.278
0.5	0.1	0.1	0.679	0.646	0.676	−3.289	−0.240
0.5	0.1	0.2	0.650	0.623	0.648	−2.769	−0.224
0.5	0.1	0.3	0.646	0.618	0.644	−2.809	−0.240
0.5	0.2	0.1	0.648	0.624	0.646	−2.398	−0.193
0.5	0.2	0.2	0.630	0.608	0.628	−2.215	−0.201
0.5	0.3	0.1	0.634	0.613	0.632	−2.067	−0.198
0.6	0.1	0.1	0.667	0.645	0.666	−2.200	−0.159
0.6	0.1	0.2	0.650	0.630	0.648	−1.969	−0.161
0.6	0.2	0.1	0.644	0.628	0.643	−1.570	−0.136
0.7	0.1	0.1	0.672	0.659	0.671	−1.305	−0.098
0.25	0.25	0.25	0.648	0.606	0.644	−4.137	−0.391

Table 10: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.1	0.1	0.1	0.901	0.947	0.924	−4.527	−2.284
0.1	0.1	0.2	0.896	0.938	0.917	−4.274	−2.114
0.1	0.1	0.3	0.890	0.931	0.911	−4.113	−1.998
0.1	0.1	0.4	0.886	0.926	0.907	−4.048	−1.939
0.1	0.1	0.5	0.883	0.923	0.904	−4.084	−1.941
0.1	0.1	0.6	0.882	0.924	0.904	−4.229	−2.010
0.1	0.1	0.7	0.887	0.932	0.910	−4.503	−2.158
0.1	0.2	0.1	0.892	0.932	0.912	−4.024	−1.998
0.1	0.2	0.2	0.887	0.925	0.906	−3.811	−1.849
0.1	0.2	0.3	0.882	0.919	0.901	−3.705	−1.762
0.1	0.2	0.4	0.878	0.916	0.898	−3.711	−1.743
0.1	0.2	0.5	0.877	0.916	0.898	−3.839	−1.797
0.1	0.2	0.6	0.881	0.922	0.903	−4.106	−1.937
0.1	0.3	0.1	0.883	0.920	0.902	−3.662	−1.779
0.1	0.3	0.2	0.878	0.913	0.897	−3.492	−1.653
0.1	0.3	0.3	0.875	0.909	0.893	−3.450	−1.603
0.1	0.3	0.4	0.873	0.909	0.892	−3.546	−1.634
0.1	0.3	0.5	0.876	0.914	0.896	−3.798	−1.762
0.1	0.4	0.1	0.875	0.910	0.893	−3.439	−1.628
0.1	0.4	0.2	0.872	0.905	0.890	−3.319	−1.531
0.1	0.4	0.3	0.870	0.903	0.888	−3.361	−1.529
0.1	0.4	0.4	0.872	0.908	0.891	−3.585	−1.636
0.1	0.5	0.1	0.869	0.903	0.887	−3.355	−1.549
0.1	0.5	0.2	0.868	0.901	0.886	−3.302	−1.491
0.1	0.5	0.3	0.869	0.904	0.888	−3.472	−1.564
0.1	0.6	0.1	0.867	0.901	0.886	−3.412	−1.546
0.1	0.6	0.2	0.869	0.904	0.888	−3.477	−1.556
0.1	0.7	0.1	0.872	0.908	0.892	−3.636	−1.635
0.2	0.1	0.1	0.888	0.924	0.906	−3.630	−1.854
0.2	0.1	0.2	0.882	0.917	0.900	−3.407	−1.702
0.2	0.1	0.3	0.878	0.911	0.895	−3.288	−1.610
0.2	0.1	0.4	0.874	0.907	0.891	−3.278	−1.583
0.2	0.1	0.5	0.873	0.907	0.891	−3.385	−1.627
0.2	0.1	0.6	0.877	0.913	0.895	−3.626	−1.754
0.2	0.2	0.1	0.879	0.911	0.895	−3.171	−1.590
0.2	0.2	0.2	0.875	0.905	0.890	−2.996	−1.462
0.2	0.2	0.3	0.871	0.900	0.886	−2.945	−1.407
0.2	0.2	0.4	0.869	0.899	0.885	−3.028	−1.432
0.2	0.2	0.5	0.872	0.904	0.889	−3.261	−1.549
0.2	0.3	0.1	0.872	0.900	0.886	−2.870	−1.400
0.2	0.3	0.2	0.868	0.896	0.883	−2.749	−1.302
0.2	0.3	0.3	0.866	0.894	0.881	−2.785	−1.296

(continued on next page)

Table 10: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.2	0.3	0.4	0.868	0.898	0.884	−2.996	−1.395
0.2	0.4	0.1	0.866	0.893	0.880	−2.727	−1.288
0.2	0.4	0.2	0.864	0.891	0.878	−2.677	−1.230
0.2	0.4	0.3	0.865	0.893	0.880	−2.841	−1.299
0.2	0.5	0.1	0.863	0.890	0.878	−2.743	−1.259
0.2	0.5	0.2	0.864	0.892	0.879	−2.809	−1.269
0.2	0.6	0.1	0.866	0.896	0.882	−2.940	−1.330
0.3	0.1	0.1	0.876	0.904	0.890	−2.842	−1.469
0.3	0.1	0.2	0.871	0.898	0.884	−2.651	−1.336
0.3	0.1	0.3	0.867	0.893	0.881	−2.583	−1.273
0.3	0.1	0.4	0.866	0.892	0.879	−2.643	−1.287
0.3	0.1	0.5	0.868	0.897	0.883	−2.848	−1.391
0.3	0.2	0.1	0.869	0.893	0.881	−2.429	−1.227
0.3	0.2	0.2	0.865	0.888	0.877	−2.300	−1.125
0.3	0.2	0.3	0.863	0.886	0.875	−2.322	−1.112
0.3	0.2	0.4	0.864	0.889	0.877	−2.513	−1.202
0.3	0.3	0.1	0.863	0.885	0.874	−2.198	−1.071
0.3	0.3	0.2	0.861	0.882	0.872	−2.145	−1.011
0.3	0.3	0.3	0.861	0.884	0.874	−2.298	−1.074
0.3	0.4	0.1	0.860	0.881	0.871	−2.152	−1.008
0.3	0.4	0.2	0.860	0.882	0.872	−2.218	−1.016
0.3	0.5	0.1	0.862	0.885	0.874	−2.309	−1.053
0.4	0.1	0.1	0.866	0.888	0.876	−2.153	−1.125
0.4	0.1	0.2	0.862	0.882	0.872	−2.003	−1.015
0.4	0.1	0.3	0.860	0.880	0.870	−2.001	−0.991
0.4	0.1	0.4	0.862	0.883	0.873	−2.161	−1.066
0.4	0.2	0.1	0.861	0.879	0.870	−1.793	−0.909
0.4	0.2	0.2	0.859	0.876	0.867	−1.728	−0.843
0.4	0.2	0.3	0.859	0.878	0.869	−1.862	−0.896
0.4	0.3	0.1	0.858	0.874	0.866	−1.652	−0.797
0.4	0.3	0.2	0.858	0.875	0.867	−1.711	−0.802
0.4	0.4	0.1	0.859	0.877	0.869	−1.749	−0.808
0.5	0.1	0.1	0.860	0.875	0.867	−1.555	−0.819
0.5	0.1	0.2	0.857	0.872	0.865	−1.463	−0.742
0.5	0.1	0.3	0.858	0.874	0.866	−1.563	−0.779
0.5	0.2	0.1	0.857	0.870	0.864	−1.266	−0.638
0.5	0.2	0.2	0.857	0.870	0.864	−1.309	−0.634
0.5	0.3	0.1	0.859	0.871	0.865	−1.265	−0.597
0.6	0.1	0.1	0.859	0.869	0.863	−1.045	−0.551
0.6	0.1	0.2	0.859	0.869	0.864	−1.050	−0.531
0.6	0.2	0.1	0.860	0.869	0.865	−0.878	−0.430
0.7	0.1	0.1	0.865	0.872	0.868	−0.637	−0.330
0.25	0.25	0.25	0.865	0.890	0.878	−2.503	−1.190

The benefits of using the mixed optimizing strategy seem evident: the loss of efficiency is much smaller when using a quasi-optimal design of this class instead of an optimal discriminating or robust design, than when a robust design is used instead of a discriminant design, or vice-versa.

5. A NOTE ON PSEUDO-CANONICAL MOMENTS OF MEASURES WITH INFINITE SUPPORT

The canonical moments are defined only for measures whose support is a subset of a closed interval. As the canonical moments are closely related with the zeros of monic orthogonal polynomials observing the recurrence relation

$$P_{m+1}(x) \;=\; \big(x - \zeta_{2m} - \zeta_{2m+1}\big)\, P_m(x) \;-\; \zeta_{2m-1}\,\zeta_{2m}\, P_{m-1}(x) \qquad \text{for} \quad m \geq 1$$

with initial conditions $P_0(x) = 1$ and $P_1(x) = x - \zeta_1$, it seems worthwhile to try to investigate some "*pseudo-canonical moments*" for measures with infinite support, using the above recurrence relation together with the recurence relation

$$P_{m+1}(x) \;=\; (A_m x + B_m)\, P_m(x) \;-\; C_m\, P_{m-1}(x)\;, \qquad m = 0, 1, 2, \ldots\;,$$

with $P_{-1}(x) = 0$ and $A_{m-1}\, A_m\, C_m > 0$, valid for any family of orthogonal polynomials.

Let us first examine the gaussian case $\mathrm{d}\mu(x) = \dfrac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\,\pi}}\,\mathrm{d}x,\ x \in \mathbb{R}$.

It is well known that the Hermite polynomials $H(x)$, recursively defined by

$$H_0(x) = 1\;; \qquad H_1(x) = x\;; \qquad H_{n+1}(x) = x\, H_n(x) - n\, H_{n-1}(x) \quad \text{for } n \geq 1$$

are orthogonal in what regards the measure μ.

Hence, in the gaussian case, the parameters ζ_m are

$$\zeta_1 = 1\,; \quad \zeta_2 = 1\,; \quad \zeta_{2m} = -\zeta_{2m+1} = (-1)^{m+1}\,\frac{m \times (m-2) \times \cdots}{(m-1) \times (m-3) \times \cdots} \quad \text{for } m \geq 2\,.$$

Using the definition $m \geq 1,\ \zeta_m = \chi^*_m\,\big(1 - \chi^*_{m-1}\big)$, we get

$$\chi^*_m = \frac{\zeta_m}{1 - \chi^*_{m-1}} \qquad (\text{with} \quad \chi^*_0 = 0)\;.$$

When $m=2$ the denominator of the previous fraction is null and therefore the gaussian distribution has only the first pseudo-canonical moment as indicated in Table 11.

Table 11: Gaussian pseudo-canonical moments, $n \leq 5$.

i	χ_i^*
1	1
2	—
3	—
4	—
5	—

Similarly, for the gamma measure with shape parameter $\alpha > 0$,

$$\mathrm{d}\mu(x) = \frac{x^{\alpha-1}\exp(-x)}{\Gamma(\alpha)}\,\mathrm{d}x\ , \qquad x \geq 0\ ,$$

which is associated with the generalized Laguerre polynomials $L^{(\alpha)}(x)$ defined by

$$L_{n+1}^{(\alpha)}(x) = (x-2n-1-\alpha)\,L_n^{(\alpha)}(x) + (n+\alpha)\,L_{n-1}^{(\alpha)}(x) \qquad \text{for} \quad n \geq 1\ ,$$

with the initial values $L_0^{(\alpha)}(x)=1,\ L_1^{(\alpha)}(x)=x-\alpha-1,$ we get

$$\begin{cases} \zeta_{2m} = -\dfrac{m+\alpha}{\zeta_{2m-1}} \\ \zeta_{2m+1} = 2m+1+\alpha-\zeta_{2m} \end{cases} .$$

Using the fact that $\zeta_0=1$ and the relation $\chi_m^* = \dfrac{\zeta_m}{1-\chi_{m-1}^*}$ (with $\chi_0^*=0$), the pseudo-canonical moments of a gamma measure with shape $\alpha=a$ up to order n are readily computed using the script

```
zeta(1) = a
zeta(2) = (1+a)/a
zeta(3) = (3+a)-zeta(2)
for j = 2:n
    zeta(2*j) = (j+a)/zeta(2*j-1)
    zeta(2*j+1) = 2*j+1+a-zeta(2*j)
end
chi(1) = zeta(1)
for j = 2:(2*n+1)
    chi(j) = zeta(j)/(1-chi(j-1))
end
```

In the table below we exhibit, as an example, the pseudo-canonical moments up to $n = 20$ for the gamma measure with shape parameter $\alpha = 3$.

Table 12: Gamma-3 pseudo-canonical moments, $n \leq 20$.

i	χ_i^*	i	χ_i^*	i	χ_i^*	i	χ_i^*
1	3	6	0.0867	11	15.4130	16	0.0323
2	0.6667	7	11.7057	12	0.0427	17	21.2785
3	22.000	8	0.0612	13	17.3558	18	0.0287
4	0.0325	9	13.4791	14	0.0368	19	23.2510
5	8.9732	10	0.0507	15	19.3125	20	0.0259

Observe that $\chi_{2n}^* \neq \frac{1}{2}$ in the case of the gaussian (while for symmetric measures with support $S \subseteq [a, b]$ we always have $\chi_{2n} = \frac{1}{2}$); or, in the case of the gamma measure, for which χ_2^* does exist, χ_2^* isn't associated with the raw moments via $\chi_2 = \dfrac{m_2 - m_1^2}{m_1(1 - m_1)}$, a relation which holds true for the canonical moments of finite support measures.

These two examples plainly show that the pseudo-canonical moments do not possess the nice properties canonical moments do satisfy in the case of measures whose support is a subset of a compact interval.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCTI and POCI/FEDER.

The authors wish to thank Prof. João Tiago Mexia (Universidade Nova de Lisboa, Portugal) for his kind encouragement during the preparation of this paper, and Prof. Roman Zmyślony (University of Zielona Góra, Poland) for his perceptive and inspiring comments at the IWMS'08 — 17$^{\text{th}}$ International Workshop on Matrices and Statistics in Tomar.

REFERENCES

[1] Anderson, T.W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem, *Ann. Math. Statist.*, **33**, 255–265.

[2] Box, G.E.P. and Draper, N.R. (1959). A basis for the selection of a response surface design, *J. Amer. Statist. Assoc.*, **54**, 622–654.

[3] Box, G.E.P. and Draper, N.R. (1963). The choice of a second order rotable design, *Biometrika*, **50**, 335–352.

[4] Dette, H. and Studden, W.J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley, New York.

[5] Fedorov, V.V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

[6] Guest, P.G. (1958). The spacing of observations in polynomial regression, *Ann. Math. Statist.*, **29**, 294–299.

[7] Henrici, P. (1977). *Applied and Computational Complex Analysis*, Vol. 2: *Special Functions, Integral Transforms, Asymptotics and Continued Fractions*, Wiley, New York.

[8] Hoel, P.G. (1958). Efficiency problems in polynomial estimation, *Ann. Math. Statist.*, **29**, 1134–1145.

[9] Kiefer, J. (1959). Optimum experimental designs, *J. Roy. Statist. Soc.*, **B21**, 273–319.

[10] Kiefer, J. (1961). Optimum designs, *Ann. Math. Statist.*, **32**, 298–325.

[11] Kiefer, J. (1962). Two more criteria equivalent to D-optimality of designs, *Ann. Math. Statist.*, **33**, 792–796.

[12] Kiefer, J. and Wolfowitz, J. (1959). Optimium designs in regression problems, *Ann. Math. Statist.*, **30**, 271–292.

[13] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems, *Canadian J. Math.*, **12**, 363–366.

[14] Kraft, O. and Schaefer, M. (1995). Exact Elfving-minimax designs for quadratic regression, *Statistics Sinica*, **5**, 475–485.

[15] Kussmaul, K. (1969). Protection against assuming the wrong degree in polynomial regression, *Technometrics*, **11**, 677–682.

[16] Lau, T.S. and Studden, W.J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments, *Ann. Statistics*, **13**, 383–394.

[17] Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley, New York.

[18] Skibinski, M. (1967). The range of the $(n+1)$-th moment for distributions on $[0, 1]$, *J. Applied Probab.*, **4**, 543–552.

[19] Skibinski, M. (1969). Some striking properties of Binomial and Beta moments, *Ann. Math. Statist.*, **40**, 1753–1764.

[20] Smith, K. (1918). On the standard deviations and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distributions of observations, *Biometrika*, **12**, 1–85.

[21] Spruill, M.C. (1990). Good designs for testing the degree of a polynomial mean, *Sankhya*, **B52**, 67–74.

[22] Stigler, S.M. (1971). Optimal experimental design for polynomial regression, *J. Amer. Statist. Assoc.*, **66**, 311–318.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in March, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract/indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.