MONOGRAFÍAS MATEMÁTICAS
"GARCÍA DE GALDEANO"

2010

N.º 36

The Pyrenees International Workshop and Summer School on Statistics, Probability and Operations Research SPO 2009

L. M. Esteban
B. Lacruz
F.J. López
P.M. Mateo
A. Pérez-Palomares
G. Sanz
C. Paroissin
(Editors)

MONOGRAFÍAS MATEMÁTICAS
"GARCÍA DE GALDEANO"

Número **36**, 2010

The Pyrenees International Workshop and Summer School on Statistics, Probability and Operations Research

SPO 2009

Jaca, Spain, September 15-18, 2009

Editors

L.M. Esteban
B. Lacruz
F.J. López
P.M. Mateo
A. Pérez-Palomares
G. Sanz
Universidad de Zaragoza, Spain

C. Paroissin
University of Pau et des Pays de l'Adour, France

The PYRENEES International Workshop and Summer School on Statistics, Probability and Operations Research : SPO 2009 : Jaca, Spain, September 15-18, 2009 / editors L.M. Esteban… [et al.]. — Zaragoza : Prensas Universitarias de Zaragoza : Seminario Matemático García de Galdeano, Universidad de Zaragoza, 2010
163 p. ; 24 cm. — (Monografías Matemáticas «García de Galdeano» ; 36)

ISBN 978-84-15031-92-5
1. Estadística matemática. 2. Probabilidades
ESTEBAN, L.M.
519.21/.22

Monografías Matemáticas García de Galdeano n.º **36**
Diciembre de 2010
Universidad de Zaragoza

Imprime: Servicio de Publicaciones. Universidad de Zaragoza

Depósito legal: Z-1536-2011
ISBN: 978-84-15031-92-5

The edition of this volume has been partially subsidized by the Vicerrectorado de Investigación de la Universidad de Zaragoza

THE PYRENEES INTERNATIONAL WORKSHOP AND SUMMER SCHOOL ON STATISTICS, PROBABILITY AND OPERATIONS RESEARCH

Jaca (Huesca), Spain, 15-18 September, 2009

SCIENTIFIC COMITTEE.

Laurent Bordes (Université de Pau et des Pays de l'Adour, France).

Wenceslao González Manteiga (Universidad de Santiago de Compostela, Spain).

Raúl Gouet (Universidad de Chile, Chile).

Fermín Mallor (Universidad Pública de Navarra, Spain).

Servet Martínez (Universidad de Chile, Chile).

José Antonio Moler (Universidad Pública de Navarra, Spain).

Domingo Morales (Universidad Miguel Hernández, Spain).

Evsey Morozov (University of Petrozavodsk, Russia).

Leandro Pardo (Universidad Complutense de Madrid, Spain).

Christian Paroissin (Université de Pau et des Pays de l'Adour, France).

Justo Puerto (Universidad de Sevilla, Spain).

Gerardo Sanz (Universidad de Zaragoza, Spain).

Joaquín Sicilia (Universidad de La Laguna, Spain).

ORGANIZING COMMITTEE.

Isolina Alberto (Universidad de Zaragoza, Spain).

Jesús Asín (Universidad de Zaragoza, Spain).

María Asunción Beamonte (Universidad de Zaragoza, Spain).

Ana Carmen Cebrián (Universidad de Zaragoza, Spain).

Luis Mariano Esteban (Universidad de Zaragoza, Spain).

Beatriz Lacruz (Universidad de Zaragoza, Spain).

David Lahoz (Universidad de Zaragoza, Spain).

Javier López (Universidad de Zaragoza, Spain).

Pedro Mateo (Universidad de Zaragoza, Spain).

Ana Pérez-Palomares (Universidad de Zaragoza, Spain).

Gerardo Sanz (Universidad de Zaragoza, Spain).

Contents

PREFACE

The Pyrenees International Workshop and Summer School on Statistics, Probability and Operations Research, SPO 2009, was held in Jaca (Spanish Pyrenees) from September 15 to September 18, 2009.

The meeting combined the structure of a workshop and a summer school with invited conferences and contributed presentations.

The school featured two advanced short courses taught by Carlos A. Coello Coello from the CINVESTAV-IPN, Mexico (A Tutorial on Multi-Objective Optimization using Metaheuristics) and Alejandro Maass from the Center for Mathematical Modelling, University of Chile, Chile (Rigidity results in cellular automata theory: probabilistic and ergodic theory approach), and a plenary conference by professor Alberto Rodríguez Casal from the University of Santiago de Compostela , Spain (Some statistical procedures for boundary estimation and image analysis). We thank them very sincerely.

In the contributed sessions, the participants introduced recent developments in Statistics, Probability and Operations Research. We also appreciated sincerely the contribution of all of them.

This volume includes extended notes of the courses developed in the conference and some of the presentations; all papers have been refereed. It is very satisfactory for us to present it to the scientific community.

We thank specially the financial support provided by Ministerio de Ciencia e Innovación (Spain) and Gobierno de Aragón. Thanks are also due to the Social Council of the University of Zaragoza, CTP (Work Community of the Pyrenees), Multicaja and University of Pau et des Pays de l'Adour. We also thank the University of Zaragoza for their financial and material support.

We wish to express our gratitude to the many colleagues (some participants at the workshop) who carefully reviewed the papers in the present volume and made many helpful suggestions for their improvement.

Special thanks are due to all members of the Scientific and Organizing committees; their generous work had a decisive influence in the success of the conference. We are also indebted to all others who helped in the organization of the conference and provided assistance to participants, in particular, Juan Marta and Daniel Sanz.

We hope that the next edition of the Pyrenees conference, SPO 2011, to be held in September 2011 will be as successful as this one.

Zaragoza, December, 2010.

The editors.

LIST OF

PARTICIPANTS

Abaurrea, Jesús
abaurrea@unizar.es

Abdul-Jalbar, Beatriz
babdul@ull.es

Ahmed, Nafidi
nafidi@hotmail.com

Alberto, Isolina
isolina@unizar.es

Alcaide, David
dalcaide@ull.es

Álvarez, Adolfo
aaapinto@est-econ.uc3m.es

Arteta, David
darteta@progenika.com

Asín, Jesus
jasin@unizar.es

Avram, Florin
florin.avram@univ-pau.fr

Balakrishnan, Narayanaswamy
bala@univmail.cis.mcmaster.ca

Barakat, Haroon Mouhamad
hbarakat2@hotmail.com

Barranco-Chamorro, Inmaculada
chamorro@us.es

Beamonte, María Asunción
asunbea@unizar.es

Bieniek, Milena
milena.bieniek@umcs.lublin.pl

Blanco, Víctor
vblanco@us.es

Blanquero, Rafael
rblanquero@us.es

Blazej, Pawel
blazej.pawel@gmail.com

Castaño, Antonia
antonia.castano@uca.es

Cebrián, Ana C.
acebrian@unizar.es

Coello, Carlos
ccoello@cs.cinvestav.mx

Deeudom, Mongkon
mongkond@kmutnb.ac.th

Dhakonlayodhin, Boonkong
bkdth@yahoo.com

Dorta, Roberto
rodorta@ull.es

Dossou-gbété, Simplice
simplice.dossou-gbete@univ-pau.fr

Esteban, Luis Mariano
lmeste@unizar.es

Gargallo, Pilar
pigarga@unizar.es

Glannamtip, Pathom
pathom@kmutnb.ac.th

Gómez-Corral, Antonio
antonio_gomez@mat.ucm.es

Jiménez, María Dolores
dolores@us.es

Jiménez, Oscar
oscarjim@et.mde.es

Jodrá Esteban, Pedro
pjodra@unizar.s

Lacruz, Beatriz
lacruz@unizar.es

Lahoz, David
davidla@unizar.es

Leisen, Fabrizio
fleisen@unav.es

Leskelä, Lasse
lasse.leskela@iki.fi

López, F. Javier
javier.lopez@unizar.es

López-Blazquez, Fernando
lopez@us.es

Maass, Alejandro
amaass@dim.uchile.cl

Madoz, Águeda
agueda.madoz@ehu.es

Magdy, El-Adll
meladll2@yahoo.com

Maiklad, Thongkam
tks@kmutnb.ac.th

Maldonado, Lina
lmguaje@unizar.es

Mallor, Fermín
mallor@unavarra.es

Martín, Nirian
nirian@estad.ucm.es

Mateo, Pedro
mateo@unizar.es

Mateos, Maider
maider.mateos@ehu.es

Miranda, Pedro
pmiranda@mat.ucm.es

Mitoi, Wansiri
wrm@kmutnb.ac.th

Moler, José Antonio
jmoler@unavarra.es

Ndongo, Mor
simplice.dossou-gbete@univ-pau.fr

Nigm, El Sayed Mahsoub
s_nigm@yahoo.com

Niwitpong, Sa-aat
snw@kmutnb.ac.th

Niwitpong, Suparat
suparat8@yahoo.com

Odiase, Justice Ighodaro
justiceodiase@yahoo.com

Osagiede, Augustine Aideyan
austaide2006@yahoo.com

Paroissin, Christian
cparoiss@univ-pau.fr

Pérez-Casany, Marta
marta.perez@upc.edu

Papageorgiou, Ioulia
ioulia@aueb.gr

Pardo, Leandro
lpardo@mat.ucm.es

Perea, Federico
perea@us.es

Phanrod, Pasakorn
pasakornp@kmutnb.ac.th

Phonyiem, Wichitra
wpy@kmutnb.ac.th

Plo, Fernando
fplo@unizar.es

Pruneda, Rosa Eva
rosa.pruneda@uclm.es

Puerto, Justo
puerto@us.es

Quesada, Jaime
jaimeque@et.mde.es

Ramos, Ana Belén
anabelenramos@us.es

Rodríguez, Alberto
alberto.rodriguez.casal@usces

Sáiz, Hugo
hsaiz@ipe.csic.es

Salami, Ali
ali.salami@univ-pau.fr

Sanz, Gerardo
gerardo@unizar.es

Sappakitkamjorn, Jeerapa
jsj@kmutnb.ac.th

Sicilia, Joaquín
jsicilia@ull.es

Sukparungsee, Saowanit
saowanits@hotmail.com

Sumontip, Kongton Janghuk
skongtun@yahoo.com

Suracherdkaiti, Wichai
wsc@kmutnb.ac.th

Szczepanska, Anna
sanna6@wp.pl

Thepsumritporn, Sthaporn
sthapornt@kmutnb.ac.th

Torrado, Nuria
nuria.torrado@uc3m.es

Turlot, Jean-Christophe
jean-christophe.turlot@univ-pau.fr

Urmeneta, Henar
henar@unavarra.es

Vivo, Juana-María
jmvivomo@um.es

Zografos, Konstantinos
kzograf@uoi.gr

Invited Papers

Monografías Matemáticas García de Galdeano **36**, 19–38 (2010)

A TUTORIAL ON MULTI-OBJECTIVE OPTIMIZATION USING METAHEURISTICS

Carlos A. Coello Coello

Abstract. This paper provides an overview of the use of metaheuristics for solving multi-objective optimization problems. The metaheuristics discussed include multi-objective evolutionary algorithms (going from the early approches to the most recent research trends in that area), multi-objective particle swarm optimizers, multi-objective artificial immune systems, multi-objective ant colony systems and multi-objective scatter search. In the final part of the paper, we provide a review of sample applications of multi-objective metaheuristics, and a discussion of some of the topics in which more research is required.

Keywords: multi-objective optimization, metaheuristics.

AMS classification: 90C29, 65K10.

§1. Introduction

A wide variety of problems in engineering and other disciplines have two or more objectives which we wish to minimize simultaneously. Such objectives are normally in conflict with each other (at least partially) and tend to be expressed in different units. These problems are called multi-objective and their solution requires a different notion of optimality that aims for the best possible trade-offs among the objectives (i.e., solutions for which no objective can be improved without worsening another one). For dealing with these problems, it is common to rely on the so-called *Pareto optimality* [92]. This definition gives rise to several compromise solutions, called the *Pareto optimal set*. The objective function values corresponding to the elements of the Pareto optimal set constitute the so-called *Pareto front*.

The algorithms for solving multi-objective optimization problems which are currently available in the mathematical programming literature [87] have a number of limitations, including the facts that some of them have a fairly limited applicability and that others need, in many cases, of problem specific information (e.g., derivatives). Additionally, some of those methods can be easily trapped in local Pareto optimal solutions when dealing with complex search spaces. This has motivated the use of alternative approaches, from which metaheuristics have gained an increasing popularity in the last few years. The main reasons for this popularity are their ease of use and their effectiveness to deal with a wide variety of problems, requiring little or no problem-specific information. Within the many types of metaheuristics currently available, evolutionary algorithms are, with no doubt, the most popular choice [19].[1]

[1] The author maintains the EMOO repository, which currently contains over 4800 bibliographic references on this topic. The EMOO repository is available at: `http://delta.cs.cinvestav.mx/~ccoello/EMOO/`

The remainder of this paper is organized as follows. In Section 2, we provide some basic concepts necessary to understand the rest of the paper. Section 3 is devoted to multi-objective evolutionary algorithms which are, the most popular multi-objective metaheuristic (MOMH) in current use. In Section 4, we talk about four more MOMHs that are relatively popular in the specialized literature. Section 5 summarizes some of the main applications of MOMHs. Two research topics that deserve further exploration are briefly discussed in Section 6. Finally, our conclusions are provided in Section 7.

§2. Basic Concepts

We are interested in solving problems of the type[2]:

$$\text{minimize } \vec{f}(\vec{x}) := [f_1(\vec{x}), f_2(\vec{x}), \ldots, f_k(\vec{x})] \tag{1}$$

subject to:

$$g_i(\vec{x}) \leq 0 \quad i = 1, 2, \ldots, m \tag{2}$$

$$h_i(\vec{x}) = 0 \quad i = 1, 2, \ldots, p \tag{3}$$

where $\vec{x} = [x_1, x_2, \ldots, x_n]^T$ is the vector of decision variables, $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, ..., k$ are the objective functions and $g_i, h_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, ..., m$, $j = 1, ..., p$ are the constraint functions of the problem.

To describe the concept of optimality in which we are interested, we will introduce next a few definitions.

Definition 1. Given two vectors $\vec{x}, \vec{y} \in \mathbb{R}^k$, we say that $\vec{x} \leq \vec{y}$ if $x_i \leq y_i$ for $i = 1, ..., k$, and that $\vec{x}$ **dominates** $\vec{y}$ (denoted by $\vec{x} \prec \vec{y}$) if $\vec{x} \leq \vec{y}$ and $\vec{x} \neq \vec{y}$.

Definition 2. We say that a vector of decision variables $\vec{x} \in \mathcal{X} \subset \mathbb{R}^n$ is **nondominated** with respect to $\mathcal{X}$, if there does not exist another $\vec{x}' \in \mathcal{X}$ such that $\vec{f}(\vec{x}') \prec \vec{f}(\vec{x})$.

Definition 3. We say that a vector of decision variables $\vec{x}^* \in \mathcal{F} \subset \mathbb{R}^n$ ($\mathcal{F}$ is the feasible region) is **Pareto-optimal** if it is nondominated with respect to $\mathcal{F}$.

Definition 4. The **Pareto Optimal Set** $\mathcal{P}^*$ is defined by:

$$\mathcal{P}^* = \{\vec{x} \in \mathcal{F} | \vec{x} \text{ is Pareto-optimal}\}$$

Definition 5. The **Pareto Front** $\mathcal{PF}^*$ is defined by:

$$\mathcal{PF}^* = \{\vec{f}(\vec{x}) \in \mathbb{R}^k | \vec{x} \in \mathcal{P}^*\}$$

Therefore, our aim is to obtain the Pareto optimal set from $\mathcal{F}$ of all the decision variable vectors that satisfy (2) and (3). It is worth indicating, however, that in practice, to obtain all the elements of the Pareto optimal set is normally undesirable and could also be impossible.

[2]Without loss of generality, we will assume only minimization problems.

Thus, our main goal when using a MOMH is to generate a good approximation of the Pareto optimal set (i.e., containing solutions that, when mapped in objective function space, are as close as possible from the true Pareto front of the problem) and having a good distribution (i.e., also in objective function space).

§3. Multi-Objective Evolutionary Algorithms

Since evolutionary algorithms are, by far, the most popular metaheuristic that has been used for solving multi-objective optimization problems, we will devote this entire section to them.

3.1. The Early Days

Evolutionary algorithm (EA) is a generic term used to denote several metaheuristics inspired on the "survival of the fittest" principle from Darwin's evolutionary theory. Their origins can be traced back to the 1960s [68, 100, 49], and have been found to be quite effective in solving a wide variety of complex search, classification and optimization problems [41].

EAs are particularly suitable for solving multi-objective optimization problems because of their capability to operate on a set of solutions (the *population*) at each iteration, which allows them to generate several trade-off solutions in a single run. They have also become popular because of their ease of use and generality (i.e., EAs are less susceptible to the shape and continuity of the Pareto front of a problem than mathematical programming techniques) [19].

It is worth indicating than traditional EAs require some modifications in order to deal with multi-objective optimization problems. The main two are the following:

1. All the nondominated solutions should be considered equally good by the selection mechanism. This means that a different notion of fitness is required for dealing with multi-objective optimization problems. The most popular mechanism to deal with this problem is called Pareto ranking and was introduced by Goldberg [59]. This approach assigns a rank to each solution based on its Pareto dominance, such that nondominated solutions are all sampled at the same rate.

2. EAs tend to converge to a single solution if run long enough, because of stochastic noise [59]. Therefore, a mechanism to maintain diversity is required. This component is known as the *density estimator*. Fitness sharing [60] was the earliest density estimator, but many others have been proposed over time, including clusters [122], entropy [47], adaptive grids [80] and crowding [31], among others.

The first actual implementation of a multi-objective evolutionary algorithm (MOEA) was David Schaffer's *Vector Evaluation Genetic Algorithm* (VEGA), which was introduced in the mid-1980s, mainly aimed for solving problems in machine learning [99].

In the period that goes from the second half of the 1980s to the first half of the 1990s, a few relatively simple and naive MOEAs were introduced. Most of them relied on aggregating functions (mostly linear) [104], lexicographic ordering [51], and target-vector approaches

[113]. Most of these MOEAs did not modify their selection mechanism or any other component, except for the definition of the fitness function. Most of these MOEAs would soon be forgotten.

As indicated before, Pareto ranking was proposed in Goldberg's famous book on genetic algorithms [59]. However, he only provided an informal description of this new selection mechanism but no actual implementation. This gave rise to several MOEAs based on Goldberg's proposal. The three most representative of the early days of MOEAs are:

1. **Nondominated Sorting Genetic Algorithm** (NSGA): This algorithm was proposed by Srinivas and Deb [103] and was the first MOEA to be published in a specialized journal (*Evolutionary Computation*). The NSGA is based on several layers of classification of the individuals as suggested by Goldberg. Before selection takes place, the population is ranked on the basis of nondominance: all nondominated individuals are classified into one category (with a dummy fitness value, which is proportional to the population size, to provide an equal reproductive potential for these individuals). To maintain the diversity of the population, these classified individuals are shared with their dummy fitness values. Then this group of classified individuals is ignored and another layer of nondominated individuals is considered. The process continues until all individuals in the population are classified. Since individuals in the first front have the maximum fitness value, they always get more copies than the rest of the population. Fitness sharing is used to distribute the population along the Pareto front of the problem.

2. **Niched-Pareto Genetic Algorithm** (NPGA): Proposed by Horn et al. [69]. It uses a tournament selection scheme based on Pareto dominance. The basic idea of the algorithm is the following: Two individuals are randomly chosen and compared against a subset from the entire population (typically, around 10% of the population). If one of them is dominated (by the individuals randomly chosen from the population) and the other is not, then the nondominated individual wins. When both competitors are either dominated or nondominated (i.e., there is a tie), the result of the tournament is decided through fitness sharing [60]. In [45], a revised version of this algorithm, called NPGA 2 was proposed. This algorithm relies on a traditional Pareto ranking approach (similar to Fonseca and Fleming's MOGA [50]), but it keeps its tournament selection scheme. Ties are solved through fitness sharing as in its predecessor. However, the niche count of the NPGA 2 is computed using individuals from the next partially filled generation using a technique called "continuously updated fitness sharing" [91].

3. **Multi-Objective Genetic Algorithm** (MOGA): Proposed by Fonseca and Fleming [50]. Here, the rank of a certain individual corresponds to the number of individuals in the current population by which it is dominated. Consider, for example, an individual x_i at generation t, which is dominated by $p_i^{(t)}$ individuals in the current generation. The rank of an individual is given by [50]:

$$\text{rank}(x_i, t) = 1 + p_i^{(t)} \tag{4}$$

   All nondominated individuals are assigned rank 1, while dominated ones are penalized according to the population density of the corresponding region of the trade-off surface.

Fitness assignment is performed in the following way [50]:

(a) Sort population according to rank.

(b) Assign fitness to individuals by interpolating from the best to the worst in the way proposed by Goldberg [59], according to some function, usually linear, but not necessarily.

(c) Average the fitnesses of individuals with the same rank, so that all of them are sampled at the same rate. This procedure keeps the global population fitness constant while maintaining appropriate selective pressure, as defined by the function used.

From these 3 algorithms, a few comparative studies undertaken during the mid and late 1990s, indicated that MOGA was the most effective and efficient approach, followed by the NPGA and by the NSGA [22, 108]. It is worth indicating that during the early days (up to the end of the 1990s), most papers would compare MOEAs without using performance measures, but only in a graphical way (plotting the Pareto fronts generated by each MOEA).

3.2. Elitist MOEAs

Towards the end of the 1990s, elitism became a standard mechanism to be provided into any MOEA. The Strengh Pareto Evolutionary Algorithm (SPEA) [118] played a key role in popularizing elitism, since it adopted an external population, and its publication in a specialized journal (the *IEEE Transactions on Evolutionary Computation*), quickly made it a landmark in the field. Consequently, many researchers started to incorporate external populations in their MOEAs, popularizing this mechanism. There are, however, also theoretical reasons for which elitism is a required mechanism in MOEAs (see [97]). Elitism consists of retaining the best solutions found during the search so that they are subject to crossover or mutation. In the context of multi-objective optimization, elitism usually (although not necessarily) refers to the use of an external population (also called secondary population) to retain the nondominated individuals found during the search. External archives can be unbounded but, mainly because of practical reasons, they are normally bounded. Another mechanism that can be used instead of an external archive is the so-called $(\mu + \lambda)$-selection in which parents compete with their children and those which are nondominated (and possibly comply with some additional criterion such as providing a better distribution of solutions) are selected for the following generation. This sort of selection is implicitly elitist, because it will retain the best half of the individuals under consideration.

With the advent of elitist MOEAs, performance measures started to become popular in the specialized literature [27, 119, 109]. It has been found, however, that some of these performance measures are not Pareto-compliant and can provide no reliable assessment [124]. There are also several benchmarks for testing new MOEAs, from which the most popular are: the Zitzler-Deb-Thiele (ZDT) test suite [119], the Deb-Thiele-Laumanns-Zitzler (DTLZ) test suite [32] and the Walkig-Fish-Group (WFG) test suite [70].

The three following approaches are representative of the elitist MOEAs in common use nowadays:

1. **Strength Pareto Evolutionary Algorithm** (SPEA): This algorithm was introduced in [118], and was conceived as a way of integrating different MOEAs. It uses an archive containing nondominated solutions previously found (the so-called external nondominated set). At each generation, nondominated individuals are copied to the external nondominated set. For each individual in this external set, a *strength* value is computed. This strength is similar to the ranking value of MOGA [50], since it is proportional to the number of solutions to which a certain individual dominates. In SPEA, the fitness of each member of the current population is computed according to the strengths of all external nondominated solutions that dominate it. The fitness assignment process of SPEA considers both closeness to the true Pareto front and even distribution of solutions at the same time. Thus, instead of using niches based on distance, Pareto dominance is used to ensure that the solutions are properly distributed along the Pareto front. Although this approach does not require a niche radius, its effectiveness relies on the size of the external nondominated set. In fact, since the external nondominated set participates in the selection process of SPEA, if its size grows too large, it might reduce the selection pressure, thus slowing down the search. Because of this, the authors decided to adopt a technique that prunes the contents of the external nondominated set so that its size remains below a certain threshold (a clustering technique called "average linkage method" [88] was used for that sake). There is a revised version of SPEA, called SPEA2, which has three main differences with respect to its predecessor [121]: (1) it incorporates a fine-grained fitness assignment strategy which takes into account for each individual the number of individuals that dominate it and the number of individuals by which it is dominated; (2) it uses a nearest neighbor density estimation technique which guides the search more efficiently, and (3) it has an enhanced archive truncation method that guarantees the preservation of boundary solutions.

2. **Pareto Archived Evolution Strategy** (PAES): This algorithm was introduced in [81]. It consists of a (1+1) evolution strategy (i.e., a single parent that generates a single offspring) in combination with a historical archive (the elitist mechanism) that records the nondominated solutions previously found. This archive is used as a reference set against which each mutated individual is being compared. An interesting aspect of this algorithm is the procedure used to maintain diversity which consists of a crowding procedure that divides objective space in a recursive manner. Each solution is placed in a certain grid location based on the values of its objectives (which are used as its "coordinates" or "geographical location"). A map of such grid is maintained, indicating the number of solutions that reside in each grid location. Since the procedure is adaptive, no extra parameters are required (except for the number of divisions of the objective space).

3. **Nondominated Sorting Genetic Algorithm II** (NSGA-II): This approach was introduced in [29, 31] as an improved version of the NSGA [103]. In the NSGA-II, for each solution one has to determine how many solutions dominate it and the set of solutions to which it dominates. The NSGA-II estimates the density of solutions surrounding a particular solution in the population by computing the average distance of two points on either side of this point along each of the objectives of the problem. This value is the so-called *crowding distance*. During selection, the NSGA-II uses a crowded-

comparison operator which takes into consideration both the nondomination rank of an individual in the population and its crowding distance (i.e., nondominated solutions are preferred over dominated solutions, but between two solutions with the same nondomination rank, the one that resides in the less crowded region is preferred). The NSGA-II combines the best parents with the best offspring obtained (i.e., a $(\mu+\lambda)$-selection), instead of using an external archive. Due to its clever mechanisms, the NSGA-II is much more efficient (computationally speaking) than its predecessor, and its performance is so good, that it has become very popular in the last few years, becoming a landmark against which other MOEAs have to be compared.

Although many other MOEAs exist (see for example [20, 101, 115]), it is not the intention of this paper to be comprehensive. The interested reader may refer to [19, 28, 105] for more information on this topic.

3.3. Current Trends in MOEAs

During some time, the use of relaxed forms of Pareto dominance became popular as a mechanism to regulate convergence of a MOEA. From these mechanisms, ϵ-dominance was, with no doubt, the most popular [85]. ϵ-dominance allows to control the granularity of the approximation of the Pareto front obtained. As a consequence, it is possible to accelerate convergence using this mechanism (if we are satisfied with a very coarse approximation of the Pareto front). Several MOEAs incorporated ϵ-dominance in their external archives (see for example [33, 65]), and there was even one MOEA fully developed around this concept (see [30]).

However, the main current research trend regarding algorithmic development is to adopt a performance measure in the selection scheme of a MOEA (hypervolume[3] has been the most popular). See for example:

- **Evolution Strategy with Probability Mutation (ESP)**: This approach uses a hypervolume-based, scaling independent, parameterless measure, to truncate overpopulated external archives [71].
- **Indicator-Based Evolutionary Algorithm (IBEA)**: This is a framework that allows any performance indicator to be incorporated into the selection mechanism of a MOEA [120]. Its authors tested it with the hypervolume and with the binary ϵ indicator.
- **S Metric Selection Evolutionary Multiobjective Algorithm (SMS-EMOA)**: This approach is based on the hypervolume performance measure [42, 9].
- **Set Preference Algorithm for Multiobjective optimization (SPAM)**: This can be seen as a generalization of IBEA which allows the use of any sort of set preference relation [123].

The use of hypervolume has some advantages, from which the main one is that it has been proved that the maximization of this performance measure is equivalent to finding the

[3] The hypervolume (also known as the S metric or the Lebesgue Measure) of a set of solutions measures the size of the portion of objective space that is dominated by those solutions collectively.

Pareto optimal set [48]. Additionally, empirical studies have shown that (for a certain number of points previously determined) the maximization of the hypervolume does indeed produce subsets of the Pareto front which are well-distributed [82, 42]. Furthermore, hypervolume measures convergence and, to a certain extent, also the spread of solutions along the Pareto front. Finally, it has been shown that hypervolume-selection is less sensitive to scalability in objective function space, which makes it promising to deal with problems having many objectives [75].

Hypervolume has, however, some problems of its own. First, the computation of this performance measure depends on a reference point, which can influence the results in a significant manner. Some people have proposed to use the worst objective function values in the current population, but this requires scaling of the objectives. Its main drawback, however, is that the best algorithms known to compute hypervolume have a polynomial complexity on the number of points used, but such complexity grows exponentially on the number of objectives. This has triggered an important amount of efforts aimed to produce more efficient algorithms to approximate the hypervolume [112, 10, 8, 7, 13].

§4. Other Metaheuristics

Several other metaheuristics have also been used as multi-objective optimizers [19, 23]. Next, we will discuss four of the most popular of them in more detail:

- **Particle Swarm Optimization**: This metaheuristic was inspired on the choreography of a bird flock which aim to find food [78, 79]. The implementation of this algorithm employs a population of particles, whose behavior is affected by either the best local (i.e., within a certain neighborhood) or the best global individual (i.e., with respect to the entire swarm). Particle swarm optimization (PSO) has been successfully used for both continuous nonlinear and discrete binary optimization [43, 44]. An important number of multi-objective versions of PSO currently exist (see for example [21, 94, 95, 46]). However, until relatively recently, most of the research had concentrated on producing new variations of existing algorithms, rather than on studying other (more interesting) topics, such as the role of the main components of a multi-objective particle swarm optimizer. Some recent research in that direction has shown that certain components that had been traditionally disregarded (e.g., the leader selection mechanism and the parameters of the flight formula) play a key role in the performance of a multi-objective particle swarm optimizer [12, 107]. There are also other interesting comparative studies aimed to identify their advantages and limitations [39]. It is expected that more research of this sort will be conducted in the next few years.

- **Artificial Immune Systems**: If considered from a computational point of view, our natural immune system can be considered a distributed intelligent system, which is able to learn and retrieve knowledge previously acquired, in order to solve several (highly complex) recognition and classification tasks [89]. These features make our immune has motivated researchers to develop computational models of our immune system which have been used for a variety of tasks, including classification, pattern recognition, and optimization [26, 89]. Several multi-objective extensions of artificial immune systems have been proposed in the specialized literature (see for example

[17, 53, 14, 52]). Also, several hybrid approaches have been proposed to solve specific tasks (see for example [2], where the authors use a multi-objective immune system hybridized with evolutionary operators and local search, in order to solve a rule extraction problem). More hybrid approaches are still to come, but until now, the high potential of multi-objective artificial immune systems in classification and pattern recognition tasks has not been fully exploited yet [117].

- **Ant Colony Optimization**: This metaheuristic was inspired on the foraging behavior of real ants. It is a distributed, stochastic search procedure based on the indirect communication of a set (called "colony") of artificial ants, which mediate using artificial pheromone trails. These pheromone trails can be seen as distributed information which is used by the ants to construct their solutions to the problem at hand. Such pheromone trails are modified during the algorithm's execution, such that they reflect the search experience acquired by the ants. This metaheuristic is intended for solving difficult (both static and dynamic) combinatorial optimization problems, in which solutions can be generated through the use of a construction procedure [36, 37]. There are several multi-objective extensions of ant colony optimization (ACO) algorithms (see for example [66, 35, 55, 1]), but as multi-objective combinatorial optimization becomes more attractive for researchers [40, 54], it is expected that more multi-objective ant colony optimization approaches (and hybrids of ACO algorithms with other metaheuristics) are proposed within the next few years.

- **Scatter Search**: This approach was originally conceived as an extension of a heuristic called surrogate constraint relaxation, which was designed for solving integer programming problems [56]. Its core idea is to adopt a series of different initializations to generate solutions. A reference set of solutions (the best found so far) is adopted, and then such solutions are "diversified" in order to generate new solutions within the neighborhood of the contents of the reference set. This sort of simple procedure is repeated until no further improvements to the contents of the reference set are detected. In the mid-1990s, some further mechanisms were added to the original scatter search algorithm, which allowed its extension to solve nonlinear, binary and permutation optimization problems [57]. These new applications triggered an important amount of research in the following years [83, 86]. Multi-objective extensions of scatter search are relatively recent, but have been steadily increasing [4, 90]. Scatter search has a lot of potential for hybrid approaches, such as memetic MOEAs [58], since it can act as a powerful local search engine for tasks such as generating missing parts of a Pareto front [98]. Because of its flexibility and ease of use, scatter search is expected to become more commonly adopted in the near future, particularly when designing hybrid MOEAs that rely heavily on good local search engines.

§5. Applications

MOEAs have been applied to a wide variety of domains (see for example [18]). However, and for the purposes of this paper, we can roughly classify the applications of MOEAs into three large groups: engineering, industrial and scientific. Some specific areas within each of

these groups are indicated next.

We will start with the engineering applications, which are, by far, the most popular in the literature. This should not be surprising, since engineering disciplines normally have problems with better understood mathematical models, which makes them more suitable for the use of MOEAs. Some sample applications of MOEAs in engineering are the following:

- Electrical engineering [111]
- Hydraulic engineering [5]
- Structural engineering [15]
- Aeronautical engineering [93]
- Robotics [106]
- Control [114]
- Telecommunications [24]
- Civil engineering [38]
- Transport engineering [61]

Now, we will provide some applications of MOEAs in industry:

- Design and manufacture [62]
- Scheduling [64]
- Management [72]

Finally, we have a variety of scientific applications of MOEAs:

- Chemistry [34]
- Physics [96]
- Medicine [125]
- Geography [6]
- Bioinformatics [3]
- Computer science [102]

Although small, this sample should give a good idea of the type of work being done with MOEAs these days. Nevertheless, many other applications exist. The interested reader should refer to the EMOO repository [16] for more information on this topic.

§6. Future Research Paths

In spite of the high volume of research done around multi-objective metaheuristics, several interesting topics remain to be explored in greater depth. Next, we briefly discuss two of them:

1. **Hybridization**: The hybridization of MOMHs with other metaheuristics and with local search mechanisms (either gradient-based or not) aimed to improve their performance is a topic that is currently being explored by many researchers (see for example [63, 67, 76, 84]), because of its high potential. Hybrid MOMHs could be viable alternatives for solving some of the great challenges of today, such as many-objective optimization problems (i.e., problems having 4 or more objective functions) [73, 74]. The use of scalarization methods combined with MOMHs is also another type of interesting hybridization that has a lot of potential for solving highly complex problems (see for example [116]).

2. **Incorporation of user's preferences**: In most real-world applications, users are not interested in the entire Pareto front, but only in a portion of it. Several mechanisms to incorporate user's preferences into a MOMH have been reported in the specialized literature (see for example [25, 77, 110, 11]), but this topic has only been scarcely explored and certainly deserves more attention.

§7. Conclusions

In this paper, we have provided a short (and highly compact) tutorial on the use of metaheuristics for solving multi-objective optimization problems. As such, this tutorial provides a very general overview of the field and is intended to serve as a quick reference for those interested in in this area. The author hopes that, in spite of favoring breadth over depth, this tutorial can be useful for those wishing to do research in multi-objective optimization using metaheuristics, since that has been the purpose of this work.

Acknowledgements

The author acknowledges support from CONACyT project no. 103570.

References

[1] A. AFSHAR, F. SHARIFI, AND M.R. JALALI. Non-dominated archiving multi-colony ant algorithm for multi-objective optimization: Application to multi-purpose reservoir operation. *Engineering Optimization*, 41(4):313–325, April 2009.

[2] J. H. ANG, K. C. TAN, AND A. A. MAMUM. An evolutionary memetic algorithm for rule extraction. *Expert Systems with Applications*, 37(2):1302–1315, March 2010.

[3] R. DILÃO, D. MURARO, M. NICOLAU, AND M. SCHOENAUER. Validation of a Morphogenesis Model of Drosophila Early Development by a Multi-objective Evolutionary Optimization

Algorithm. In C. Pizzuti, M. D. Ritchie, and M. Giacobini, editors, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBIO'2009)*, pages 176–190. Springer, Lecture Notes in Computer Science, Vol. 5483, Tübingen, Germany, 2009. ISBN 978-3-642-01183-2.

[4] R. P. BEAUSOLEIL. "MOSS" multiobjective scatter search applied to non-linear multiple criteria optimization. *European Journal of Operational Research*, 169(2):426–449, March 2006.

[5] E. G. BEKELE AND J. W. NICKLOW. Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology*, 341(3-4):165–176, August 2007.

[6] D. A. BENNETT, N. XIAO, AND M. P. ARMSTRONG. Exploring the Geographic Consequences of Public Policies Using Evolutionary Algorithms. *Annals of the Association of American Geographers*, 94(4):827–847, 2004.

[7] N. BEUME. S-Metric Calculation by Considering Dominated Hypervolume as Klee's Measure Problem. *Evolutionary Computation*, 17(4):477–492, Winter 2009.

[8] N. BEUME, C. M. FONSECA, M. LOPEZ-IBANEZ, L. PAQUETE, AND J. VAHRENHOLD. On the Complexity of Computing the Hypervolume Indicator. *IEEE Transactions on Evolutionary Computation*, 13(5):1075–1082, October 2009.

[9] N. BEUME, B. NAUJOKS, AND M. EMMERICH. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 16 September 2007.

[10] L. BRADSTREET, L. WHILE, AND L. BARONE. A Fast Incremental Hypervolume Algorithm. *IEEE Transactions on Evolutionary Computation*, 12(6):714–723, December 2008.

[11] J. BRANKE AND K. DEB. Integrating User Preferences into Evolutionary Multi-Objective Optimization. In Yaochu Jin, editor, *Knowledge Incorporation in Evolutionary Computation*, pages 461–477. Springer, Berlin Heidelberg, 2005. ISBN 3-540-22902-7.

[12] J. BRANKE AND S. MOSTAGHIM. About Selecting the Personal Best in Multi-Objective Particle Swarm Optimization. In T. P. Runarsson et al., editors, *Parallel Problem Solving from Nature - PPSN IX, 9th International Conference*, pages 523–532. Springer. Lecture Notes in Computer Science Vol. 4193, Reykjavik, Iceland, September 2006.

[13] K. BRINGMANN AND T. FRIEDRICH. Approximating the Least Hypervolume Contributor: NP-Hard in General, But Fast in Practice. In M. Ehrgott, C. M. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, editors, *Evolutionary Multi-Criterion Optimization. 5th International Conference, EMO 2009*, pages 6–20. Springer. Lecture Notes in Computer Science Vol. 5467, Nantes, France, April 2009.

[14] F. CAMPELO, F. G. GUIMARÃES, AND H. IGARASHI. Overview of Artificial Immune Systems for Multi-Objective Optimization. In S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, editors, *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007*, pages 937–951, Matshushima, Japan, March 2007. Springer. Lecture Notes in Computer Science Vol. 4403.

[15] T.Y. CHEN AND H.C. CHEN. Mixed-discrete structural optimization using a rank-niche evolution strategy. *Engineering Optimization*, 41(1):39–58, January 2009.

[16] C. A. COELLO COELLO. The EMOO repository: a resource for doing research in evolutionary multiobjective optimization. *IEEE Computational Intelligence Magazine*, 1(1):37–45, February 2006.

[17] C. A. COELLO COELLO AND N. CRUZ CORTÉS. Solving Multiobjective Optimization Problems using an Artificial Immune System. *Genetic Programming and Evolvable Machines*, 6(2):163–190, June 2005.

[18] C. A. COELLO COELLO AND G. B. LAMONT, editors. *Applications of Multi-Objective Evolutionary Algorithms*. World Scientific, Singapore, 2004. ISBN 981-256-106-4.

[19] C. A. COELLO COELLO, G. B. LAMONT, AND D. A. VAN VELDHUIZEN. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer, New York, second edition, September 2007. ISBN 978-0-387-33254-3.

[20] C. A. COELLO COELLO AND G. TOSCANO PULIDO. Multiobjective Optimization using a Micro-Genetic Algorithm. In L. Spector et al., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2001)*, pages 274–282, San Francisco, California, 2001. Morgan Kaufmann Publishers.

[21] C. A. COELLO COELLO, G. TOSCANO PULIDO, AND M. SALAZAR LECHUGA. Handling Multiple Objectives With Particle Swarm Optimization. *IEEE Transactions on Evolutionary Computation*, 8(3):256–279, June 2004.

[22] C. A. COELLO COELLO. *An Empirical Study of Evolutionary Techniques for Multiobjective Optimization in Engineering Design*. PhD thesis, Department of Computer Science, Tulane University, New Orleans, LA, April 1996.

[23] D. CORNE, M. DORIGO, AND F. GLOVER, editors. *New Ideas in Optimization*. McGraw-Hill, London, 1999.

[24] J. CRICHIGNO AND B. BARÁN. Multiobjective Multicast Routing Algorithm. In J. Neuman de Souza, P. Dini, and P. Lorenz, editors, *Telecommunications and Networking. 11th International Conference on Telecommunications (ICT'2004)*, pages 1029–1034. Springer, Lecture Notes in Computer Science, Vol. 3124, Fortaleza, Brazil, August 1-6 2004. ISBN 978-3-540-22571-3.

[25] D. CVETKOVIĆ AND I. C. PARMEE. Preferences and their Application in Evolutionary Multiobjective Optimisation. *IEEE Transactions on Evolutionary Computation*, 6(1):42–57, February 2002.

[26] D. DASGUPTA, editor. *Artificial Immune Systems and Their Applications*. Springer-Verlag, Berlin, 1999.

[27] K. DEB. Multi-Objective Genetic Algorithms: Problem Difficulties and Construction of Test Problems. *Evolutionary Computation*, 7(3):205–230, Fall 1999.

[28] K. DEB. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001. ISBN 0-471-87339-X.

[29] K. DEB, S. AGRAWAL, A. PRATAP, AND T. MEYARIVAN. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In M. Schoenauer et al., editors, *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pages 849–858, Paris, France, 2000. Springer. Lecture Notes in Computer Science No. 1917.

[30] K. DEB, M. MOHAN, AND S. MISHRA. Evaluating the ϵ-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions. *Evolutionary Computation*, 13(4):501–525, Winter 2005.

[31] K. DEB, A. PRATAP, S. AGARWAL, AND T. MEYARIVAN. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA–II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[32] K. DEB, L. THIELE, M. LAUMANNS, AND E. ZITZLER. Scalable Test Problems for Evolutionary Multiobjective Optimization. In A. Abraham, L. Jain, and R. Goldberg, editors, *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, pages 105–145. Springer, USA, 2005.

[33] V. DEVIREDDY AND P. REED. Efficient and Reliable Evolutionary Multiobjective Optimization Using ε-Dominance Archiving and Adaptive Population Sizing. In K. Deb et al., editor, *Genetic and Evolutionary Computation–GECCO 2004. Proceedings of the Genetic and Evolutionary Computation Conference. Part II*, pages 390–391, Seattle, Washington, USA, June 2004. Springer-Verlag, Lecture Notes in Computer Science Vol. 3103.

[34] A. DIETZ, A. AGUILAR-LASSERRE, C. AZZARO-PANTEL, L. PIBOULEAU, AND S. DOMENECH. A fuzzy multiobjective algorithm for multiproduct batch plant: Application to protein production. *Computers & Chemical Engineering*, 32(1-2):292–306, January-February 2008.

[35] K.F. DOERNER, W.J. GUTJAHR, R.F. HARTL, C. STRAUSS, AND C. STUMMER. Pareto ant colony optimization with ILP preprocessing in multiobjective portfolio selection. *European Journal of Operational Research*, 171(3):830–841, June 2006.

[36] M. DORIGO AND G. DI CARO. The Ant Colony Optimization Meta-Heuristic. In David Corne, Marco Dorigo, and Fred Glover, editors, *New Ideas in Optimization*, pages 11–32, London, 1999. McGraw-Hill.

[37] M. DORIGO AND T. STÜTZLE. *Ant Colony Optimization*. The MIT Press, 2004. ISBN 0-262-04219-3.

[38] L. DRIDI, M. PARIZEAU, A. MAILHOT, AND J.-P. VILLENEUVE. Using evolutionary optimization techniques for scheduling water pipe renewal considering a shor planning horizon. *Computer-Aided Civil and Infrastructure*, 23(8):625–635, November 2008.

[39] J. J. DURILLO, J. GARCÍA-NIETO, A. J. NEBRO, C. A. COELLO COELLO, F. LUNA, AND E. ALBA. Multi-Objective Particle Swarm Optimizers: An Experimental Comparison. In M. Ehrgott, C. M. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, editors, *Evolutionary Multi-Criterion Optimization. 5th International Conference, EMO 2009*, pages 495–509. Springer. Lecture Notes in Computer Science Vol. 5467, Nantes, France, April 2009.

[40] M. EHRGOTT AND X. GANDIBLEUX. Multiobjective Combinatorial Optimization—Theory, Methodology, and Applications. In M. Ehrgott and X. Gandibleux, editors, *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*, pages 369–444. Kluwer Academic Publishers, Boston, 2002.

[41] A.E. EIBEN AND J.E. SMITH. *Introduction to Evolutionary Computing*. Springer, Berlin, 2003. ISBN 3-540-40184-9.

[42] M. EMMERICH, N. BEUME, AND B. NAUJOKS. An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, pages 62–76, Guanajuato, México, March 2005. Springer. Lecture Notes in Computer Science Vol. 3410.

[43] A. P. ENGELBRECHT. *Computational Intelligence: An Introduction*. John Wiley & Sons, 2003. ISBN 0-47084-870-7.

[44] A. P. ENGELBRECHT. *Fundamentals of Computational Swarm Intelligence*. John Wiley & Sons, Ltd, 2005. ISBN 0-470-09191-6.

[45] M. ERICKSON, A. MAYER, AND J. HORN. The Niched Pareto Genetic Algorithm 2 Applied to the Design of Groundwater Remediation Systems. In E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, editors, *First International Conference on Evolutionary Multi-Criterion Optimization*, pages 681–695. Springer-Verlag. Lecture Notes in Computer Science No. 1993, 2001.

[46] S.-K. FAN AND J.-M. CHANG. A parallel particle swarm optimization algorithm for multi-objective optimization problems. *Engineering Optimization*, 41(7):673–697, July 2009.

[47] A. FARHANG-MEHR AND S. AZARM. Diversity Assessment of Pareto Optimal Solution Sets: An Entropy Approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey, May 2002. IEEE Service Center.

[48] M. FLEISCHER. The Measure of Pareto Optima. Applications to Multi-objective Metaheuristics. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 519–533, Faro, Portugal, April 2003. Springer. Lecture Notes in Computer Science. Volume 2632.

[49] L. J. FOGEL. *Artificial Intelligence through Simulated Evolution*. John Wiley, New York, 1966.

[50] C. M. FONSECA AND P. J. FLEMING. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416–423, San Mateo, California, 1993. University of Illinois at Urbana-Champaign, Morgan Kauffman Publishers.

[51] M. P. FOURMAN. Compaction of Symbolic Layout using Genetic Algorithms. In *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, pages 141–153. Lawrence Erlbaum, 1985.

[52] F. FRESCHI, C. A. COELLO COELLO, AND M. REPETTO. Multiobjective Optimization and Artificial Immune Systems: A Review. In H. Mo, editor, *Handbook of Research on Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies*, pages 1–21. Medical Information Science Reference, Hershey, New York, 2009. ISBN 978-1-60566-310-4.

[53] F. FRESCHI AND M. REPETTO. VIS: an artificial immune network for multi-objective optimization. *Engineering Optimization*, 38(8):975–996, December 2006.

[54] X. GANDIBLEUX AND M. EHRGOTT. 1984-2004 – 20 Years of Multiobjective Metaheuristics. But What About the Solution of Combinatorial Problems with Multiple Objectives? In C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, pages 33–46, Guanajuato, México, March 2005. Springer. Lecture Notes in Computer Science Vol. 3410.

[55] C. GARCÍA-MARTÍNEZ, O. CORDÓN, AND F. HERRERA. A taxonomy and an empirical analysis of multiple objective ant colony optimization algorithms for the bi-criteria TSP. *European Journal of Operational Research*, 180(1):116–148, July 2007.

[56] F. GLOVER. Heuristics for integer programming using surrogate constraints. *Decision Sciences*, 8:156–166, 1977.

[57] F. GLOVER. Tabu search for nonlinear and parametric optimization (with links to genetic algorithms). *Discrete Applied Mathematics*, 49:231–255, 1994.

[58] C.-K. GOH, Y.-S. ONG, AND K. C. TAN, editors. *Multi-Objective Memetic Algorithms*. Springer, Berlin, Germany, 2009. ISBN 978-3-540-88050-9.

[59] D. E. GOLDBERG. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[60] D. E. GOLDBERG AND J. RICHARDSON. Genetic algorithm with sharing for multimodal function optimization. In J. J. Grefenstette, editor, *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*, pages 41–49, Hillsdale, New Jersey, 1987. Lawrence Erlbaum.

[61] M. M. GOLIAS, M. BOILE, AND S. THEOFANIS. Berth scheduling by customer service differentiation: A multi-objective approach. *Transportation Research Part E - Logistics and Transportation Review*, 45(6):878–892, November 2009.

[62] L. GRANDINETTI, F. GUERRIERO, G. LEPERA, AND M. MANCINI. A niched genetic algorithm to solve a pollutant emission reduction problem in the manufacturing industry: A case study. *Computers & Operations Research*, 34(7):2191–2214, July 2007.

[63] J. GROBLER AND A. P. ENGELBRECHT. Hybridizing PSO and DE for improved vector evaluated multi-objective optimization. In *2009 IEEE Congress on Evolutionary Computation (CEC'2009)*, pages 1255–1262, Trondheim, Norway, May 2009. IEEE Press.

[64] J. GROBLER, A. P. ENGELBRECHT, AND V. S. S. YADAVALLI. Multi-Objective DE and PSO Strategies for Production Scheduling. In *2008 Congress on Evolutionary Computation (CEC'2008)*, pages 1154–1161, Hong Kong, June 2008. IEEE Service Center.

[65] C. GROSAN. Improving the Performance of Evolutionary Algorithms for the Multiobjective 0/1 Knapsack Problem using ε-dominance. In *2004 Congress on Evolutionary Computation (CEC'2004)*, volume 2, pages 1958–1963, Portland, Oregon, USA, June 2004. IEEE Service Center.

[66] M. GUNTSCH AND M. MIDDENDORF. Solving Multi-criteria Optimization Problems with Population-Based ACO. In C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, and L. Thiele, editors, *Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003*, pages 464–478, Faro, Portugal, April 2003. Springer. Lecture Notes in Computer Science. Volume 2632.

[67] K. HARADA, K. IKEDA, AND S. KOBAYASHI. Hybridizing of Genetic Algorithm and Local Search in Multiobjective Function Optimization: Recommendation of GA then LS. In M. Keijzer et al., editor, *2006 Genetic and Evolutionary Computation Conference (GECCO'2006)*, volume 1, pages 667–674, Seattle, Washington, USA, July 2006. ACM Press. ISBN 1-59593-186-4.

[68] J. H. HOLLAND. Concerning efficient adaptive systems. In M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, editors, *Self-Organizing Systems—1962*, pages 215–230. Spartan Books, Washington, D.C., 1962.

[69] J. HORN, N. NAFPLIOTIS, AND D. E. GOLDBERG. A Niched Pareto Genetic Algorithm for Multiobjective Optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, volume 1, pages 82–87, Piscataway, New Jersey, June 1994. IEEE Service Center.

[70] S. HUBAND, P. HINGSTON, L. BARONE, AND L. WHILE. A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, October 2006.

[71] S. HUBAND, P. HINGSTON, L. WHITE, AND L. BARONE. An Evolution Strategy with Probabilistic Mutation for Multi-Objective Optimisation. In *Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003)*, volume 3, pages 2284–2291, Canberra, Australia, December 2003. IEEE Press.

[72] A. K. HUTZSCHENREUTER, P. A. N. BOSMAN, AND H. LA POUTRÉ. Evolutionary Multiobjective Optimization for Dynamic Hospital Resource Management. In M. Ehrgott, C. M.

Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, editors, *Evolutionary Multi-Criterion Optimization. 5th International Conference, EMO 2009*, pages 320–334. Springer. Lecture Notes in Computer Science Vol. 5467, Nantes, France, April 2009.

[73] H. ISHIBUCHI, Y. HITOTSUYANAGI, AND Y. NOJIMA. Scalability of Multiobjective Genetic Local Search to Many-Objective Problems: Knapsack Problem Case Studies. In *2008 Congress on Evolutionary Computation (CEC'2008)*, pages 3587–3594, Hong Kong, June 2008. IEEE Service Center.

[74] H. ISHIBUCHI, N. TSUKAMOTO, AND Y. NOJIMA. Evolutionary many-objective optimization: A short review. In *2008 Congress on Evolutionary Computation (CEC'2008)*, pages 2424–2431, Hong Kong, June 2008. IEEE Service Center.

[75] H. ISHIBUCHI, N. TSUKAMOTO, Y. SAKANE, AND Y. NOJIMA. Hypervolume Approximation Using Achievement Scalarizing Functions for Evolutionary Many-Objective Optimization. In *2009 IEEE Congress on Evolutionary Computation (CEC'2009)*, pages 530–537, Trondheim, Norway, May 2009. IEEE Press.

[76] S. JEONG, S. HASEGAWA, K. SHIMOYAMA, AND S. OBAYASHI. Development and Investigation of Efficient GA/PSO-Hybrid Algorithm Applicable to Real-World Design Optimization. *IEEE Computational Intelligence Magazine*, 4(3):36–44, August 2009.

[77] Y. JIN AND B. SENDHOFF. Incorporation of Fuzzy Preferences into Evolutionary Multiobjective Optimization. In W.B. Langdon et al., editors, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2002)*, page 683, San Francisco, California, July 2002. Morgan Kaufmann Publishers.

[78] J. KENNEDY AND R. C. EBERHART. Particle Swarm Optimization. In *Proceedings of the 1995 IEEE International Conference on Neural Networks*, pages 1942–1948, Piscataway, New Jersey, 1995. IEEE Service Center.

[79] J. KENNEDY AND R. C. EBERHART. *Swarm Intelligence*. Morgan Kaufmann Publishers, San Francisco, California, 2001.

[80] J. KNOWLES AND D. CORNE. Properties of an Adaptive Archiving Algorithm for Storing Nondominated Vectors. *IEEE Transactions on Evolutionary Computation*, 7(2):100–116, April 2003.

[81] J. D. KNOWLES AND D. W. CORNE. Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation*, 8(2):149–172, 2000.

[82] J. D. KNOWLES, D. W. CORNE, AND M. FLEISCHER. Bounded Archiving using the Lebesgue Measure. In *Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003)*, volume 4, pages 2490–2497, Canberra, Australia, December 2003. IEEE Press.

[83] M. LAGUNA AND R. MARTÍ. *Scatter Search : Methodology and Implementations in C*. Kluwer Academic Publishers, 2003. ISBN 1-402-07376-3.

[84] A. LARA, G. SANCHEZ, C. A. COELLO COELLO, AND O. SCHÜTZE. HCS: A New Local Search Strategy for Memetic Multi-Objective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 14(1):112–132, February 2010.

[85] M. LAUMANNS, L. THIELE, K. DEB, AND E. ZITZLER. Combining Convergence and Diversity in Evolutionary Multi-objective Optimization. *Evolutionary Computation*, 10(3):263–282, Fall 2002.

[86] R. MARTÍ. Scatter Search–Wellsprings and Challenges. *European Journal of Operational Research*, 169:351–358, 2006.

[87] K. M. MIETTINEN. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, Massachusetts, 1999.

[88] J.N. MORSE. Reducing the size of the nondominated set: Pruning by clustering. *Computers and Operations Research*, 7(1–2):55–66, 1980.

[89] L. N. DE CASTRO AND J. TIMMIS. *An Introduction to Artificial Immune Systems: A New Computational Intelligence Paradigm*. Springer, London, 2002. ISBN 1-85233-594-7.

[90] A. J. NEBRO, F. LUNA, E. ALBA, B. DORRONSORO, J. J. DURILLO, AND A. BEHAM. AbYSS: Adapting Scatter Search to Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation*, 12(4):439–457, August 2008.

[91] C. K. OEI, D. E. GOLDBERG, AND S.-J. CHANG. Tournament Selection, Niching, and the Preservation of Diversity. Technical Report 91011, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, December 1991.

[92] V. PARETO. *Cours D'Economie Politique*, volume I and II. F. Rouge, Lausanne, 1896.

[93] S. RAIAGOPAL AND R. GANGULI. Conceptual design of UAV using Kriging based multi-objective genetic algorithm. *Aeronautical Journal*, 112(1137):653–662, November 2008.

[94] M. REYES SIERRA AND C. A. COELLO COELLO. Improving PSO-Based Multi-objective Optimization Using Crowding, Mutation and ϵ-Dominance. In C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization. Third International Conference, EMO 2005*, pages 505–519, Guanajuato, México, March 2005. Springer. Lecture Notes in Computer Science Vol. 3410.

[95] M. REYES-SIERRA AND C. A. COELLO COELLO. Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art. *International Journal of Computational Intelligence Research*, 2(3):287–308, 2006.

[96] V. ROMERO-GARCIA, J. V. SANCHEZ-PEREZ, L. M. GARCIA-RAFFI, J. M. HERRERO, S. GARCIA-NIETO, AND X. BLASCO. High Optimization processes for increasing the attenuation properties of acoustic metamaterials by means of creation of defects. *Applied Physics Letters*, 93(22), December 2008.

[97] G. RUDOLPH AND A. AGAPIE. Convergence Properties of Some Multi-Objective Evolutionary Algorithms. In *Proceedings of the 2000 Conference on Evolutionary Computation*, volume 2, pages 1010–1016, Piscataway, New Jersey, July 2000. IEEE Press.

[98] L. V. SANTANA-QUINTERO, N. RAMÍREZ, AND C. COELLO COELLO. A Multi-objective Particle Swarm Optimizer Hybridized with Scatter Search. In A. Gelbukh and C. A. Reyes-Garcia, editors, *MICAI 2006: Advances in Artificial Intelligence, 5th Mexican International Conference on Artificial Intelligence*, pages 294–304. Springer, Lecture Notes in Artificial Intelligence Vol. 4293, Apizaco, Mexico, November 2006.

[99] J. D. SCHAFFER. Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. In *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, pages 93–100. Lawrence Erlbaum, 1985.

[100] H.-P. SCHWEFEL. Kybernetische evolution als strategie der experimentellen forschung in der strömungstechnik. Dipl.-Ing. thesis, 1965. (in German).

[101] M. SEFRIOUI AND J. PERIAUX. Nash Genetic Algorithms: examples and applications. In *2000 Congress on Evolutionary Computation*, volume 1, pages 509–516, San Diego, California, July 2000. IEEE Service Center.

[102] A. J. SOTO, R. L. CECCHINI, G. E. VAZQUEZ, AND I. PONZONI. Multi-Objective Feature Selection in QSAR Using a Machine Learning Approach. *QSAR & Combinatorial Science*, 28(11-12):1509–1523, December 2009.

[103] N. SRINIVAS AND K. DEB. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3):221–248, Fall 1994.

[104] G. SYSWERDA AND J. PALMUCCI. The Application of Genetic Algorithms to Resource Scheduling. In R. K. Belew and L. B. Booker, editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 502–508, San Mateo, California, 1991. Morgan Kaufmann.

[105] K.C. TAN, E.F. KHOR, AND T.H. LEE. *Multiobjective Evolutionary Algorithms and Applications*. Springer-Verlag, London, 2005. ISBN 1-85233-836-9.

[106] J. TEO, L. D. NERI, M. H. NGUYEN, AND H. A. ABBASS. Walking with EMO: Multi-Objective Robotics for Evolving Two, Four, and Six–Legged Locomotion. In L. T. Bui and S. Alam, editors, *Multi-Objective Optimization in Computational Intelligence: Theory and Practice*, pages 300–332. Information Science Reference, Hershey, PA, USA, 2008. ISBN 978-1-59904-498-9.

[107] G. TOSCANO-PULIDO, C. A. COELLO COELLO, AND L. V. SANTANA-QUINTERO. EMOPSO: A Multi-Objective Particle Swarm Optimizer with Emphasis on Efficiency. In S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, editors, *Evolutionary Multi-Criterion Optimization, 4th International Conference, EMO 2007*, pages 272–285, Matshushima, Japan, March 2007. Springer. Lecture Notes in Computer Science Vol. 4403.

[108] D. A. VAN VELDHUIZEN. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, May 1999.

[109] D. A. VAN VELDHUIZEN AND GARY B. LAMONT. On Measuring Multiobjective Evolutionary Algorithm Performance. In *2000 Congress on Evolutionary Computation*, volume 1, pages 204–211, Piscataway, New Jersey, July 2000. IEEE Service Center.

[110] J. WANG AND J. P. TERPENNY. Interactive Preference Incorporation in Evolutionary Engineering Design. In Yaochu Jin, editor, *Knowledge Incorporation in Evolutionary Computation*, pages 525–543. Springer, Berlin Heidelberg, 2005. ISBN 3-540-22902-7.

[111] L. WANG AND C. SINGH. Stochastic combined heat and power dispatch based on multi-objective particle swarm optimization. *International Journal of Electrical Power & Energy Systems*, 30(3):226–234, March 2008.

[112] L. WHILE, P. HINGSTON, L. BARONE, AND S. HUBAND. A Faster Algorithm for Calculating Hypervolume. *IEEE Transactions on Evolutionary Computation*, 10(1):29–38, February 2006.

[113] P. B. WIENKE, C. LUCASIUS, AND G. KATEMAN. Multicriteria target optimization of analytical procedures using a genetic algorithm. *Analytica Chimica Acta*, 265(2):211–225, 1992.

[114] L. XU, B. ZHU, AND E. D. GOODMAN. An Improved MOCC with Feedback Control Structure Based on Preference. In *2009 ACM SIGEVO Summit on Genetic and Evolutionary Computation (GEC'2009)*, pages 651–656, Shanghai, China, June 12-14 2009. ACM Press. ISBN 978-1-60558-326-6.

[115] S. Y. ZENG, L. S. KANG, AND L. X. DING. An Orthogonal Multi-objective Evolutionary Algorithm for Multi-objective Optimization Problems with Constraints. *Evolutionary Computation*, 12(1):77–98, Spring 2004.

[116] Q. ZHANG AND H. LI. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, December 2007.

[117] X. ZHANG, B. LU, S. GOU, AND L. JIAO. Immune Multiobjective Optimization Algorithm Using Unsupervised Feature Selection. In F. Rothlauf et al., editor, *Applications of Evolutionary Computing. EvoWorkshops 2006: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, and EvoSTOC*, pages 484–494, Budapest, Hungary, April 2006. Springer, Lecture Notes in Computer Science Vol. 3907.

[118] E. ZITZLER, K. DEB, AND L. THIELE. Comparison of Multiobjective Evolutionary Algorithms on Test Functions of Different Difficulty. In A. S. Wu, editor, *Proceedings of the 1999 Genetic and Evolutionary Computation Conference. Workshop Program*, pages 121–122, Orlando, Florida, July 1999.

[119] E. ZITZLER, K. DEB, AND L. THIELE. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, Summer 2000.

[120] E. ZITZLER AND S. KÜNZLI. Indicator-based Selection in Multiobjective Search. In X. Yao et al., editor, *Parallel Problem Solving from Nature - PPSN VIII*, pages 832–842, Birmingham, UK, September 2004. Springer-Verlag. Lecture Notes in Computer Science Vol. 3242.

[121] E. ZITZLER, M. LAUMANNS, AND L. THIELE. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. In K. Giannakoglou, D. Tsahalis, J. Periaux, P. Papailou, and T. Fogarty, editors, *EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, pages 95–100, Athens, Greece, 2002.

[122] E. ZITZLER AND L. THIELE. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, November 1999.

[123] E. ZITZLER, L. THIELE, AND J. BADER. SPAM: Set Preference Algorithm for Multiobjective Optimization. In Günter Rudolph, Thomas Jansen, Simon Lucas, Carlo Poloni, and Nicola Beume, editors, *Parallel Problem Solving from Nature–PPSN X*, pages 847–858. Springer. Lecture Notes in Computer Science Vol. 5199, Dortmund, Germany, September 2008.

[124] E. ZITZLER, L. THIELE, M. LAUMANNS, C. M. FONSECA, AND V. GRUNERT DA FONSECA. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, April 2003.

[125] V. ZOETE, A. GROSDIDIER, AND O. MICHELIN. Docking, Virtual High Troughput Screening and in Silico Fragment-Based Drug Design. *Journal of Cellular and Molecular Medicine*, 13(2):238–248, February 2009.

Carlos A. Coello Coello
CINVESTAV-IPN
Departamento de Computación
Av. IPN No. 2508
Col. San Pedro Zacatenco
México, D.F. 07360, México
ccoello@cs.cinvestav.mx

Monografías Matemáticas García de Galdeano **36**, 39–56 (2010)

RIGIDITY RESULTS IN CELLULAR AUTOMATA THEORY: PROBABILISTIC AND ERGODIC THEORY APPROACH

Alejandro Maass

Abstract. This article contains the talk given at the Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 09. Its objective is to review the main results and its extensions concerning the existence of invariant stationary probability measures under a one-dimensional *algebraic* cellular automaton. We present two historical axes of this question and the techniques used to solve them or produce relevant intermediate results. Both make appear strong *rigidity* phenomena, i.e. the unique solution is the uniform Bernoulli product measure. The first axe is the ergodic theory approach where we impose some natural conditions on the entropy and ergodicity of the system to get the result. This approach follows ideas by Rudolph [18] and Host [9] in the classical problem called $(\times 2, \times 3)$ in the circle posed by Hillel Furstenberg at the end of the 60'. Then we present a purely probabilistic approach. We study the convergence of the Cesàro mean of the iterates by an algebraic cellular automaton of a translation invariant probability measure. Assuming some natural correlation properties (this class includes Markov and Gibbs probability measures of full topological support) one proves the limit exists and it is the uniform Bernoulli product measure.

Keywords: Cellular automata, Entropy, Dynamical systems

AMS classification: 37B15

Contents

§1. Introduction

The study of the dynamics of cellular automata is often associated to its capacity to model complex systems using very simple local interactions in the phase space. From the probabilistic point of view one possible interpretation of such complex behavior is the richness of the space of invariant probability measures and the existence of limit measures for the iterates of interesting classes of probability measures by cellular automata. Such initial conditions can be thought as the law that we use to produce random configurations in the phase space of the cellular automata. Both questions represent major challenges of the theory of cellular automata and few results exist even in the one-dimensional setting.

A remarkable situation appears when considering classes of algebraic cellular automata. That is, when in addition, the underlying phase space has an algebraic structure compatible with the cellular automata. In this context, it has been observed that even if the sets of invariant probability measures are rich, under very natural conditions the unique invariant measure is the maximal entropy one. From the perspective of iterating probability measures by cellular automata of algebraic origin these kinds of results lead to think that good candidates to be limits of such sequences of iterations are the same class of measures. In fact, such limits rarely exist being the Cesàro means of such sequences the good candidates to converge. Indeed, it has been observed by D. Lind in his pioneer work [12] that the Cesàro mean of the iterates of a Bernoulli measure on $\{0,1\}^{\mathbb{Z}}$ by the cellular automaton $F = \sigma^{-1} + \sigma$ converges to the uniform product measure $(\frac{1}{2}, \frac{1}{2})^{\mathbb{Z}}$. In this example σ is the shift map and $\{0,1\}^{\mathbb{Z}}$ is the product Abelian group with addition modulo two componentwise. Since Lind's results, this *rigid* phenomenon has being proved to be very general in the algebraic context and seems to be extensible to the context of positively expansive or expansive cellular automata. The purpose of these notes is to present what in the opinion of the author are the most illustrative results concerning the described phenomena in the case of algebraic cellular automata acting on a fullshift $A^{\mathbb{Z}}$ that at the same time is a compact Abelian group.

The notes are organized as follows. Section 2 is devoted to a minimum of background in symbolic dynamics and ergodic theory. In section 3 we present a basic example to illuminate the problems in a relevant *study case*. The rigidity phenomenon that appears in the set of invariant measures is presented in section 4. There we show that different ergodicity conditions together with some entropy conditions imply that the Haar measure is the unique invariant measure for the shift and the cellular automaton simultaneously. Section 5 is devoted to the study of the convergence of the Cesàro means of the iterates of nice classes of probability measures by algebraic cellular automata.

§2. Preliminaries and Background

2.1. Symbolic Dynamics (in dimension 1)

In this section we summarize the main background in Symbolic Dynamics that we will need in the article. For a more detailed exposition we suggest the book by D. Lind and B. Marcus [11].

Let A be a finite set or *alphabet*. Following this last nomenclature its elements are also called *letters* or *symbols*. Denote by A^* the set of finite sequences or *words* $w = w_0...w_{n-1} \in A^n$

with letters in A including the empty word ε. By $|w|$ we mean the length of $w \in A^*$ and $|\varepsilon| = 0$.
Let $X = A^{\mathbb{Z}}$ be the set of two-sided sequences

$$\mathbf{x} = (x_i)_{i\in\mathbb{Z}} = (\dots x_{-i} \dots x_0 \dots x_i \dots)$$

with symbols in A. Analogously one defines $X = A^{\mathbb{N}}$ to be the set of one-sided sequences in A. Both are called *full-shifts*. For simplicity we restrict our attention to the two-sided case. The space X is compact for the product topology and metrizable. A classical distance is given by:

$$d(\mathbf{x}, \mathbf{y}) = 2^{-\inf\{|i| \,:\, i\in\mathbb{Z},\ x_i\neq y_i\}},$$

for any $\mathbf{x}, \mathbf{y} \in X$, i.e. two points are close if they coincide in big windows near the origin. For $\mathbf{x} \in X$ and $i \leq j$ in $\mathbb{Z}$ or $\mathbf{x} = x_0 \dots x_n \in A^*$ and $i \leq j$ in $\{0, \dots, n\}$, we denote by $x[i, j] = x_i \dots x_j$ the finite word in $\mathbf{x}$ between coordinates i and j. Given $w \in A^*$ and $i \in \mathbb{Z}$, the cylinder set starting in coordinate i with word w is $[w]_i = \{\mathbf{x} \in X : x[i, i+|w|-1] = w\}$.
A natural dynamical system on X is the *shift map* $\sigma : X \to X$, where $\sigma(\mathbf{x}) = (x_{i+1})_{i\in\mathbb{Z}}$. This map is a homeomorphism of X. If we need to distinguish a shift map according to its alphabet we denote it by σ_A.
We call $Y \subseteq X$ a *subshift* if it is closed (for the product topology) and $\sigma(Y) = Y$ (invariant for the shift map). A simple example is given by the orbit closure of a point $\mathbf{x}$ in X, i.e. $\overline{\{\sigma^n(\mathbf{x}) \,:\, n \in \mathbb{Z}\}}$. A special class of subshifts are *shifts of finite type*; they are inspired in Markov chains in probability theory. A subshift $Y \subseteq X$ is a subshift of finite type if there is a finite subset $\mathcal{W}$ of words in A of a given length L such that for any $\mathbf{y} \in Y$ and $i \in \mathbb{Z}$,

$$y_i \dots y_{i+L-1} \notin \mathcal{W}.$$

Example 2.1. Let $A = \{0, 1\}$ and consider $\mathcal{W} = \{11\}$. The subshift of finite type Y defined by $\mathcal{W}$ as described before consists of two-sided infinite sequences in A do not containing two consecutive ones.

The *language* of a subshift $Y \subseteq X$ is given by

$$\mathcal{L}(Y) = \{y[i, j] : \mathbf{y} \in Y,\ i, j \in \mathbb{Z},\ i \leq j\}.$$

One saids that Y is *mixing* if for any $u, v \in \mathcal{L}(Y)$ there is $N \in \mathbb{N}$ such that for any $n > N$, $[u]_0 \cap \sigma^{-n}[v]_0 \neq \emptyset$. In other words, there is a word $w \in A^n$ such that $uwv \in \mathcal{L}(Y)$.
A second kind of important dynamics are given by continuous and shift commuting maps of a subshift $Y \subseteq X$. A map $F : Y \to Y$ is called a *block-map* if F is continuous (for the product topology) and $F \circ \sigma = \sigma \circ F$. They are called block maps since Hedlund's result [8] states that there is a *local map* $f : A^{m+a+1} \to A$, where $a, m \in \mathbb{N}$ (a is called anticipation and m memory), such that $\forall\, i \in \mathbb{Z}, \forall\, \mathbf{y} \in Y$

$$F(\mathbf{y})_i = f(y_{i-m}, \dots, y_{i+a}).$$

We also use f to indicate the action of the local rule on words of length greater than or equal to $m + a + 1$. That is, for $w = w_0...w_n \in A^*$ with $|w| \geq m + a + 1$, we put

$$f(w) = f(w_0, ..., w_{m+a}) f(w_1, ..., w_{m+a+1}) ... f(w_{n-(m+a)}, ..., w_n).$$

If $m = 0$ or $a = 0$ one says that F is *one-sided*.

2.2. Cellular automata and invariant measures

Let $Y \subseteq A^{\mathbb{Z}}$ be a *mixing* shift of finite type and $F : Y \to Y$ a block-map. Then F is called a *cellular automaton* (CA). Typical examples correspond to $Y = A^{\mathbb{Z}}$ (a full-shift). Analogously one defines cellular automata acting on the set of one-sided subshifts $Y \subseteq A^{\mathbb{N}}$ but this case is not considered in these notes.
Let $Y \subseteq A^{\mathbb{Z}}$ be a subshift and $F : Y \to Y$ a block-map. A probability measure μ defined on the Borel σ-algebra of Y (we simply say "on Y") is *F-invariant* if for any Borel set B of Y

$$F\mu(B) := \mu(F^{-1}(B)) = \mu(B).$$

If $F = \sigma$ (the shift map on Y), the measure is said to be *stationary* or *shift invariant*. An invariant measure is *ergodic* if invariant Borel sets have measure 0 or 1. We observe that if μ is shift invariant, since F commutes wit the shift, then $F^n\mu$ is also shift invariant, where F^n is the n-th iterate of F.
In this paper we study the convergence of the *Cesàro mean* $\mathcal{M}_\mu^N(F) = \frac{1}{N}\sum_{n=0}^{N-1} F^n\mu$. If this limit exists as $N \to \infty$, we denote it by $\mathcal{M}_\mu(F)$. If $Y = A^{\mathbb{Z}}$ a main role will be played by the uniform product or Bernoulli measure $\lambda_A^{\mathbb{Z}}$ of $A^{\mathbb{Z}}$, where λ_A is the equidistributed probability measure on A.
The following classes of CA on $A^{\mathbb{Z}}$ are relevant for these notes:

1) *Linear CA.* Let $(A, +)$ be a finite Abelian group. This structure naturally extends to $A^{\mathbb{Z}}$ by componentwise operations, so $(A^{\mathbb{Z}}, +)$ is also an Abelian group (to simplify notations we also denote the operation by $+$). Observe that $(A^{\mathbb{Z}}, +)$ is a compact Abelian group and the uniform Bernoulli measure is its *Haar measure*. It is characterized as the unique probability measure μ such that $\mu(\chi) = \int_{A^{\mathbb{Z}}} \chi d\mu = 0$ for every non-trivial character $\chi \in \widehat{A^{\mathbb{Z}}}$, i.e. for $\chi \neq 1$.
A cellular automaton $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is said to be *linear* if for any $\mathbf{x} \in A^{\mathbb{Z}}$

$$F(\mathbf{x}) = \sum_{i=1}^{l} k_i \sigma^{n_i}(\mathbf{x})$$

where $n_1, \ldots, n_l, k_1, \ldots, k_l \in \mathbb{Z}$. In terms of the local rule this means:

$$F(\mathbf{x})_j = \sum_{i=1}^{l} k_i x_{j+n_i} \ .$$

2) *Permutative CA.* The CA $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is said to be *right permutative* if for every $w \in A^{m+a}$ the map $f(w, \cdot) : A \to A$ defined by $f(w, \cdot)(\alpha) = f(w\alpha)$ is one-to-one. This implies that for every $w \in A^{m+a}$ the map $f_w : A^{m+a} \to A^{m+a}$ given by $f_w(w') = f(ww')$ is also one-to-one. Analogously we define *left permutative* and *bipermutative* CA.

2.3. Entropy

A classical measure of complexity of the dynamics of a (classically surjective) CA F on a subshift $Y \subseteq A^{\mathbb{Z}}$ with respect to an F invariant measure μ is the *Shannon entropy* or just

the *entropy*. It is defined as follows: for any $N \in \mathbb{N}$ let $\alpha_{N,\infty}$ be the σ-algebra given by $\bigvee_{n\geq 1} F^{-n}\alpha_N$, where $\alpha_N = \{\mathbf{y}[-N,N] : \mathbf{y} \in Y\}$ is a partition of Y, and put:

$$h_\mu(F) = -\lim_{N\to\infty} \sum_{C\in\alpha_N} \int_Y 1_C(\mathbf{y}) \log\left(\mathbb{E}(1_C|\alpha_{N,\infty})(\mathbf{y})\right) d\mu(\mathbf{y}) \,.$$

A probablity measure of *maximal entropy* (for the CA) is one for which:

$$h_\mu(F) = \sup_\nu h_\nu(F)$$

where the $\sup$ is taken over all F invariant probability measures on Y.

§3. Main Questions

3.1. Questions and comments

Let $F : Y \to Y$ be a surjective cellular automaton on a mixing shift of finite type Y. In this article we assume $Y = A^{\mathbb{Z}}$. Here we set the three main questions of this theory.

Question 1: Study the set of *invariant measures* of F and in addition of the joint action of F and σ. That is, find probability measures μ on Y such that for any Borel set B and integers $n \in \mathbb{N}, m \in \mathbb{Z}$

$$F^n\mu(B) := \mu(F^{-n}B) = \mu(B)$$

or

$$F^n \circ \sigma^m\mu(B) := \mu(F^{-n} \circ \sigma^{-m}B) = \mu(B).$$

A natural invariant measure for F is the one of *maximal entropy* for the shift map. In fact, it holds that F is surjective if and only if the maximal entropy measure is also F-invariant [4]. Depending on the subshift Y and the dynamical properties of F it is possible to construct other invariant measures (see for example [26]). Nevertheless, in some cases strong rigidities appear, that is, this is the unique shift and F invariant probability measure.

Question 2: Given a shift invariant probability measure μ on Y study if the limit of the sequence $(F^n\mu : n \in \mathbb{N})$ exists (in the weak topology). We remark that every weak limit of a subsequence is invariant for F and the shift. It is also interesting the convergence when $N \to \infty$ of the Cesàro mean

$$\mathcal{M}_\mu^N(F) = \frac{1}{N}\sum_{n=0}^{N-1} F^n\mu \,.$$

One says F *asymptotically randomizes* μ if the limit of the Cesàro mean converges to the maximal entropy measure.

Question 3: Find conditions to ensure the maximal entropy measure is the unique solution to Questions 1 and 2.

Comments:
— In relation with Question 1 the type of solutions we look for are like the $(\times 2, \times 3)$ Furstenberg's problem in $\mathbb{R}/\mathbb{Z}$ [18]: F (or σ) is ergodic and σ (resp. F) has positive entropy for the invariant measure. While ergodicity of one transformation can be changed for a weaker condition the positivity of the entropy cannot be dropped for the moment. Proofs strongly rely on entropy formulas. These conditions already appear in Rudolph's or Host's solutions to $(\times 2, \times 3)$ problem and all recent improvements (see [9, 18]).

— In relation with Question 2 in the linear case there are two points of view. One is to consider measures μ of increasing complexity in correlations: Markov, Gibbs, other chain connected measures; then represent them as "independent processes" and prove that the limit of the Cesàro mean converges to the uniform Bernoulli product measure on $A^{\mathbb{Z}}$ [6]. The other one is motivated by Lind's work [12] and uses harmonic analysis. The idea is to define a class of *mixing* measures such that the Cesàro mean of the iterates of any of them converges.

– From Glasner and Weiss results in topological dynamics (see [7]) one gets that either the CA map F is *almost equicontinuous* or *sensitive to initial conditions*, and in the last class most interesting known examples (and in fact coming from Nasu's reductions [17]) are *expansive* or *positively expansive maps*. In the equicontinuous case or systems with equicontinuous points, orbits tend to be periodic and invariant measures can be more or less described but are not nice. Moreover, in this case the limits of the Cesàro means we are considering always converge [2]. If the CA are positively expansive they are *conjugate with shifts of finite type* (see [1, 3]), so we have two commuting shifts of finite type with the same maximal entropy measure. In this last case there can still exist an *equicontinuous direction* so invariant measures are as in the previous case.

– Good examples: the classes of positively expansive or expansive CA without equicontinuous directions have not been described. The main examples with these features correspond to *algebraic maps*, in particular linear CA. This is why most of the results concerning Questions 1, 2 and 3 are concentrated on these maps.

3.2. Basic example: addition modulo 2 or *Ledrappier's three dot problem*

Let $X = \{0,1\}^{\mathbb{Z}}$ and see X as an Abelian group with coordinatewise addition modulo 2. Let $F : X \to X$ be given by $F = id + \sigma$, where σ is the shift map on X. That is, $F(\mathbf{x})_i = x_i + x_{i+1}$. Remark that it is a 2-to-1 surjective map.

– In relation to Question 1: Natural invariant measures for F and the shift map simultaneously are the uniform Bernoulli product measure $\lambda = (1/2, 1/2)^{\mathbb{Z}}$ and measures supported on periodic orbits for F and the shift, but other invariant measures of algebraic origin has been described (see [26]).

– In relation to Question 2: In general the limit does not exist. It follows from a good understanding of the Pascal triangle modulo 2.
We give a brief argument in the Bernoulli case. Let $\mu = (\pi_0, \pi_1)^{\mathbb{Z}}$ be a Bernoulli non-uniform

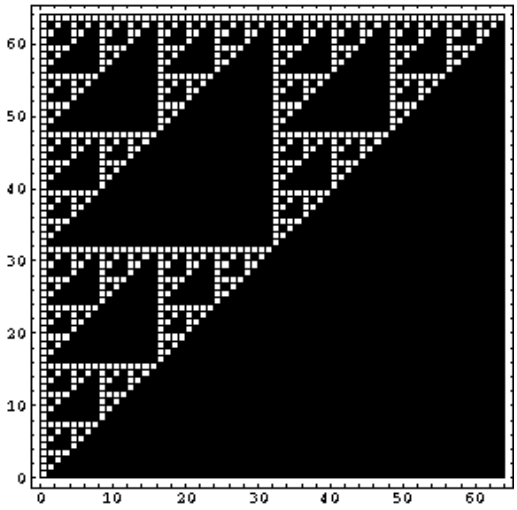


Figure 1: The iteration of a ultimately periodic configuration by the addition modulo 2 CA. One observes the Pascal triangle modulo 2 behind it.

product measure on X with $\pi_0 = \mu(x_i = 0)$, $\pi_1 = \mu(x_i = 1)$. A simple induction yields to:

$$\mu\left(\sum_{i\in I} x_i = a\right) = \frac{1}{2}\left(1 + (-1)^a(\pi_0 - \pi_1)^{\#I}\right)$$

where I is a finite subset of $\mathbb{N}$. Thus,

$$F^n\mu([a]_0) = \mu\left(\sum_{k\in I(n)} x_k = a\right) = \frac{1}{2}\left(1 + (-1)^a(\pi_0 - \pi_1)^{\#I(n)}\right)$$

where $I(n) = \{0 \leq k \leq n \ : \ \binom{n}{k} = 1 \bmod 2\}$.
If $a = 0$, for the subsequence along $n = 2^m$ the limit exists and is equal to $\pi_0^2 + \pi_1^2$ and for the subsequence along $n = 2^m - 1$ the limit is $\frac{1}{2}$.
But the Cesàro mean converges:

$$\mathcal{M}_\mu^N(F)([a]_0) = \frac{1}{N}\sum_{n=0}^{N-1} F^n\mu([a]_0) = \frac{1}{2} + \frac{(-1)^a}{2}\frac{1}{N}\sum_{n=0}^{N-1}(\pi_0 - \pi_1)^{\#I(n)}$$

since $\lim_{N\to\infty} \frac{\{0\leq n<N\,:\,\#I(n)\geq \alpha \log\log N\}}{N} = 1$ for some $\alpha \in (0,1/2)$ (a non trivial consequence of Lucas' lemma, see [6]) then the limit is $\frac{1}{2}$. Also, using similar arguments, one proves that $\mathcal{M}_\mu^N(F)([a_0 \dots a_{l-1}]_0)$ converges to $\frac{1}{2^l}$ as $N \to \infty$. This was observed by D. Lind in 84 for $F = \sigma^{-1} + \sigma$ [12].

This result reinforces the idea that the uniform Bernoulli product measure $\lambda = (1/2, 1/2)^{\mathbb{Z}}$ must be the unique invariant measure of F and σ verifying *some conditions to be determined*. The following sections will go deeper on these conditions and proofs will be drafted for this particular but relevant example.

§4. Ergodic approach: results on invariant measures for a CA and the shift simultaneously

4.1. A theorem for the basic example

The *model* theorem in the theory concerns the basic example. Most of the existing generalizations start from this example changing the precise local map, the cardinality of the alphabet or the algebraic structure of $A^{\mathbb{Z}}$. In the rest of this section $\{0,1\}^{\mathbb{Z}}$ is seen as a product Abelian group where addition modulo 2 is applied componentwise.

Theorem 4.1 (Basic Theorem [10])**.** *Let $F : \{0,1\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ be the CA given by $F = id + \sigma$. If μ is an F and σ invariant probability measure on $\{0,1\}^{\mathbb{Z}}$ with $h_\mu(F) > 0$ and is ergodic for σ then $\mu = \lambda = (1/2, 1/2)^{\mathbb{Z}}$.*

Sketch of the proof. Let μ be a simultaneously invariant probability measure for F and the shift σ. Put $X = \{0,1\}^{\mathbb{Z}}$. We describe the main steps of the proof, the difficulty is not homogeneous but each step only requires a little computation:

1) Let $\mathcal{B}$ be the product σ-algebra of X and $\mathcal{B}_1 = F^{-1}\mathcal{B}$. For μ–a.e. $\mathbf{x} \in X$ define $\mu_{\mathbf{x}}(\cdot) = \mathbb{E}(\cdot|\mathcal{B}_1)(\mathbf{x})$. This measure is concentrated on $\{\mathbf{x}, \mathbf{x}+\mathbf{1}\}$, where $\mathbf{1} = \dots 1111111 \dots \in X$. Also, $\sigma\mu_{\mathbf{x}} = \mu_{\sigma\mathbf{x}}$.
2) Define $\phi(\mathbf{x}) = \mu_{\mathbf{x}}(\{\mathbf{x}+\mathbf{1}\})$. Then

$$\phi \circ \sigma(\mathbf{x}) = \mu_{\sigma\mathbf{x}}(\{\sigma\mathbf{x}+\mathbf{1}\}) = \sigma\mu_{\mathbf{x}}(\{\sigma\mathbf{x}+\mathbf{1}\}) = \mu_{\mathbf{x}}(\{\mathbf{x}+\mathbf{1}\}) = \phi(\mathbf{x})$$

3) The **ergodicity** of μ with respect to σ implies that ϕ is constant μ–a.e., thus also $F\mu$-a.e. This implies that

$$\phi \circ F = \phi \circ \sigma = \phi,\ \mu - \text{a.e.} \qquad (*)$$

4) Define $E = \{\mathbf{x} \in X\ :\ \phi(\mathbf{x}) > 0\}$.

– If $B \subseteq X$ is the *good set* of measure one where $(*)$ is satisfied then $\mu\{\mathbf{x} \in E : \mathbf{x}+\mathbf{1} \in B\} = \mu(E)$;

– so, $\phi(\mathbf{x}+\mathbf{1}) = \phi(F(\mathbf{x}+\mathbf{1})) = \phi(F(\mathbf{x})) = \phi(\mathbf{x})$ for μ-a.e. $\mathbf{x}$ in E;

– that is, $\mu_{\mathbf{x}}(\{\mathbf{x}\}) = \mu_{\mathbf{x}}(\{\mathbf{x}+\mathbf{1}\}) = \frac{1}{2}$ for μ-a.e. $\mathbf{x}$ in E.

5) E is σ-invariant by (*), then by **ergodicity** $\mu(E) = 0$ or $\mu(E) = 1$.

6) **Entropy formula:** let $\alpha = \{[0]_0, [1]_0\}$. Therefore, using standard computations in the entropy theory of dynamical systems one deduces (see for instance [19]):

$$h_\mu(F) = -\sum_{a\in\{0,1\}} \int_X 1_{[a]_0}(\mathbf{x}) \log\left(\mathbb{E}(1_{[a]_0}|\mathcal{B}_1)(\mathbf{x})\right) d\mu(\mathbf{x})$$

Observe that when $\mathbf{x} \in [a]_0$ then $\mu_{\mathbf{x}}([a]_0) = \mu(\{\mathbf{x}\})$ since $\mathbf{x} + \mathbf{1} \notin [a]_0$ for $a = 0, 1$. Then

$$h_\mu(F) = -\int_X \log(\mu_{\mathbf{x}}(\{\mathbf{x}\}))d\mu(\mathbf{x}) \ .$$

7) Conclusion:

– If $h_\mu(F) > 0$ then from 6) one deduces that $\mu(E) > 0$. Therefore, from 5) (**ergodicity**) follows that $\mu(E) = 1$;

– this last fact implies by 4) that: $\mu_{\mathbf{x}}(\{\mathbf{x}\}) = \mu_{\mathbf{x}}(\{\mathbf{x} + \mathbf{1}\}) = \frac{1}{2}$ for μ-a.e. $\mathbf{x} \in X$;

– concluding by 6) that $h_\mu(F) = \log(2)$. Since $\lambda = (1/2, 1/2)^{\mathbb{Z}}$ is the unique maximal entropy measure for F (or similarly is the unique stationary probability measure on X verifying last equality) then $\mu = \lambda$.

□

4.2. Some generalizations

As commented before, several generalizations can be expected. The next two are of different nature. The first one consists just in changing $\{0, 1\}^{\mathbb{Z}}$ by $\mathbb{Z}_p = \{0, \ldots, p - 1\}^{\mathbb{Z}}$ where p is a prime number. Its proof is essentially copying the one given before. In the second one, the ergodicity condition is weaker, then we need to add a condition on the σ-algebra of invariant sets. Here, we do not change the local rules, but all reasonable linear extensions follows also directly.

Theorem 4.2 (Host, Maass, Martínez, [10])**.** *Let $F : \mathbb{Z}_p^{\mathbb{Z}} \to \mathbb{Z}_p^{\mathbb{Z}}$ be given by $F = id + \sigma$. Let μ be an F and σ invariant probability measure on $\mathbb{Z}_p^{\mathbb{Z}}$. If $h_\mu(F) > 0$ and μ is ergodic for the shift then μ is the uniform product measure $(1/p, \ldots, 1/p)^{\mathbb{Z}}$.*

We need an additional concept. One says that an invariant probability measure μ for a CA $F : Y \to Y$ and the shift map on Y is (F, σ)-ergodic if any Borel set B in Y that is invariant for the joint action of such maps has measure 0 or 1, i.e., if $\mu(F^{-n}\sigma^{-m}B\Delta B) = 0$ for any $n \in \mathbb{N}, m \in \mathbb{Z}$ then $\mu(B) = 0$ or $\mu(B) = 1$. We denote by $\mathcal{I}_\mu(F) = \{B \in \mathcal{B}(Y) : \mu(F^{-1}B\Delta B) = 0\}$, the set of invariant Borel sets for F with respect to μ.

Theorem 4.3 (Host, Maass, Martínez, [10]). *Let $F : \mathbb{Z}_p^{\mathbb{Z}} \to \mathbb{Z}_p^{\mathbb{Z}}$ be given by $F = id + \sigma$. Let μ be an F and σ invariant probability measure on $\mathbb{Z}_p^{\mathbb{Z}}$. If $h_\mu(F) > 0$, μ is (F, σ)-ergodic and $\mathcal{I}_\mu(\sigma) = \mathcal{I}_\mu(\sigma^{p(p-1)})$, then μ is the uniform product measure $(1/p, \ldots, 1/p)^{\mathbb{Z}}$.*

The next generalizations try to extract from previous theorems those properties that seems to be the main objects behind this class of results. We say a CA $F : Y \to Y$ is *algebraic* if Y (in addition to be a mixing shift of finite type) is a compact Abelian topological group and F and the shift are endomorphisms of such group. Here, the role of the uniform product measure is played by the Haar measure of the compact Abelian group (the unique probability measure that is invariant by translation by elements of the group). An invariant measure for F is said to be *totally ergodic* if it is ergodic for all powers of F. This is a very strong ergodicity condition that allows frequently in this theory to jump over the difficult obstacles. It implies the technical hypothesis about invariant σ-algebras in Theorem 4.3, but this last is a much more refined condition.

Theorem 4.4 (Pivato, [23]). *Let $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be an algebraic bipermutative CA. If μ is a totally ergodic invariant probability measure for σ, $h_\mu(F) > 0$ and $Ker(F)$ has no shift invariant subgroups, then μ is the Haar measure of $A^{\mathbb{Z}}$.*

The most general extension of Theorem 4.3 not using total ergodicity is the following.

Theorem 4.5 (Sablik, [25]). *Let $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be an algebraic bipermutative CA and Σ be an F and σ invariant closed subgroup of $A^{\mathbb{Z}}$. Fix $k \in \mathbb{N}$ such that any prime divisor of $|A|$ divides k. Let μ be an F and σ invariant probability measure on $A^{\mathbb{Z}}$ with supp$(\mu) \subseteq \Sigma$ such that:*
– μ is (F, σ)-ergodic;
– $h_\mu(F) > 0$;
– $\mathcal{I}_\mu(\sigma) = \mathcal{I}_\mu(\sigma^{kp_1})$, where p_1 is the smallest common period of the elements in $Ker(F)$;
– any finite shift invariant subgroup of $\cup_{n\in\mathbb{N}} Ker(F^n) \cap \Sigma$ is dense in Σ.
Then μ is the Haar measure of Σ.

These theorems have an analogous result in a much more general algebraic context. We do not give all the details of this theory since they escape from the context of these notes. The main point is that, instead of considering the actions of a CA F and the shift map on a mixing shift of finite type Y, one considers d commuting actions on a 0-dimensional set. In our context $d = 2$, the commuting actions are F and σ and the 0-dimensional space is Y.

Theorem 4.6 (Einsiedler, [5]). *Let α be an algebraic $\mathbb{Z}^d$-action of a compact 0-dimensional Abelian group verifying some algebraic conditions. Let μ be an invariant (for the complete action) probability measure. Then, if the action has positive entropy in one direction and the measure is totally ergodic for the action then it is the Haar measure of the group.*

Remark 4.1. From last theorems it is possible to deduce the same kind of results for some classes of positively expansive and expansive CA actions on a fullshift, *a priori* not algebraic (see [10]).

§5. Probabilistic approach: results on the convergence of Cesàro means

In this section we will present the main results concerning the convergence of Cesàro means of the iterates of a probability measure by algebraic cellular automata. In [12] D. Lind proposes an harmonic analysis point of view to study the convergence of the Cesàro means of the iterates of a Bernoulli product measure by the CA $\sigma^{-1}+\sigma$ on $\{0,1\}^{\mathbb{Z}}$ seen as a product Abelian group. This technique cannot work alone, it needs a fine combinatorial analysis of the Pascal triangle modulo 2. The extension of Lind's pioneer results to other classes of initial probability measures and other types of algebraic cellular automata was considered in [13] and was deepened in [6]. In these works the main example is $id+\sigma$ in $\mathbb{Z}_p^{\mathbb{Z}}$ where p is a prime number. The main issue was that Bernoulli measures were changed by probability measures with complete connections and summable decay of correlations, class that includes Markov and Gibbs measures for example. The harmonic analysis ideas used by Lind were not used here, the technical part was to represent such general class of measures using independent processes via regeneration idea. The harmonic analysis reappeared with the works of M. Pivato and R. Yassawi ([20, 21]). They put into the concept of *harmonically mixing* measures (introduced by them) the main properties observed in the classes of measures considered in [6], and the dynamical properties of the algebraic CA considered (that essentially comes from Pascal triangle) gave rise to the concept of *diffusivity*, giving an abstract formalization to previous results. The complexification of the classes of algebraic CA considered was achieved in several further works [22, 23, 24, 14, 15, 16]. For simplicity we will not describe such results in this review. The main statements there propose the same kind of results as in the basic (but fundamental) cases up to some natural and necessary technical conditions.

5.1. Harmonic analysis point of view

Let $(A,+)$ be a finite Abelian group and fix $F:A^{\mathbb{Z}}\to A^{\mathbb{Z}}$ an algebraic CA.
A character $\chi: A\to\mathbb{T}^1$ in $\widehat{A^{\mathbb{Z}}}$, where $\mathbb{T}^1$ is the one-dimensional torus, is given by $\chi=\bigotimes_{k\in\mathbb{Z}}\chi_k$ where χ_k are characters of A and $\chi_k=1$ for all but finitely many terms in this product. The rank of the character χ, rank(χ), is the number of non trivial characters χ_k in $\bigotimes_{k\in\mathbb{Z}}\chi_k$.

The Haar or uniform Bernoulli measure λ on $A^{\mathbb{Z}}$ is characterized by

$$\lambda(\chi)=\int_{A^{\mathbb{Z}}}\chi d\lambda=0\quad\forall\chi\not\equiv 1\,. \tag{1}$$

Definition 5.1 (Pivato, Yassawi, [20])**.** A probability measure μ on $A^{\mathbb{Z}}$ is *harmonically mixing* if $\forall\varepsilon>0\ \exists N(\varepsilon)>0$ such that $\forall\chi\in\widehat{A^{\mathbb{Z}}}$:

$$\text{rank}(\chi)>N(\varepsilon)\Rightarrow|\mu(\chi)|=\left|\int_{A^{\mathbb{Z}}}\chi d\mu\right|<\varepsilon\,.$$

If $A=\mathbb{Z}_p$, then a Markov probability measure with strictly positive transitions is harmonically mixing.

Definition 5.2 (Pivato, Yassawi, [20])**.** • The block map $F:A^{\mathbb{Z}}\to A^{\mathbb{Z}}$ is *diffusive* if

$$\forall\chi\not\equiv 1:\lim_{n\to\infty}\text{rank}\,[\chi\circ F^n]=\infty.$$

• F is *diffusive in density* if there exists $J \subseteq \mathbb{N}$ of density 1 such that

$$\lim_{\substack{n\to\infty \\ n\in J}} \text{rank } [\chi \circ F^n] = \infty.$$

The following theorem can be considered as a consolidation of results in [20, 21, 6]. Nevertheless, historically this form appeared first in [21].

Theorem 5.1. *Let $(A,+)$ be a finite Abelian group and $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be an algebraic CA with local map given by $f(x_{i-m} \dots x_{i+a}) = \sum_{k=-m}^{a} f_k(x_{i+k})$, where $f_{-m}, \dots, f_a$ are commuting automorphisms of the group $(A,+)$ and at least two of them are non-trivial. Then F is diffusive in density and for any harmonically mixing probability measure μ on $A^{\mathbb{Z}}$:*

$$\mathcal{M}_\mu(F) = \lim_{N\to\infty} \mathcal{M}_\mu^N(F) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} F^n\mu = \lambda.$$

Sketch of the proof. The diffusitivy of F comes from the combinatorial structure of the associated generalized Pascal triangle and we do not give it here. The consequence on the convergence of the Cesàro means follows directly from the combination of diffusivity of F and the harmonically mixing condition of μ. Indeed, let χ be a non-trivial character in $\widehat{A}^{\mathbb{Z}}$ and $N \in \mathbb{N}$. A simple computation yields to,

$$\int_{A^{\mathbb{Z}}} \chi d\mathcal{M}_\mu^N(F) = \frac{1}{N} \sum_{n=0}^{N-1} \int_{A^{\mathbb{Z}}} \chi \circ F^n d\mu \,.$$

Assume F is diffusive in density and consider $J \subseteq \mathbb{N}$ with density 1 such that $\lim_{\substack{n\to\infty \\ n\in J}} \text{rank } [\chi \circ F^n] = \infty$ and put $J_N = J \cap \{0, \dots, N-1\}$. One gets,

$$\int_{A^{\mathbb{Z}}} \chi d\mathcal{M}_\mu^N(F) = \frac{1}{N} \sum_{n\in J_N} \int_{A^{\mathbb{Z}}} \chi \circ F^n d\mu + \frac{1}{N} \sum_{n\in J_N^c} \int_{A^{\mathbb{Z}}} \chi \circ F^n d\mu \,.$$

Thus,

$$\left| \int_{A^{\mathbb{Z}}} \chi d\mathcal{M}_\mu^N(F) \right| \leq \left| \frac{1}{N} \sum_{n\in J_N} \int_{A^{\mathbb{Z}}} \chi \circ F^n d\mu \right| + \frac{|J_N^c|}{N} \,.$$

Since μ is harmonically mixing, given $\epsilon > 0$ there is $N(\epsilon) \in \mathbb{N}$ such that for any $n \geq N(\epsilon)$ in J_N,

$$\left| \int_{A^{\mathbb{Z}}} \chi \circ F^n d\mu \right| \leq \epsilon \,.$$

One concludes, taking the limit as $N \to \infty$ that:

$$\lim_{N\to\infty} \int_{A^{\mathbb{Z}}} \chi d\mathcal{M}_\mu^N(F) \leq \epsilon \,.$$

By (1), this implies that $\mathcal{M}_\mu^N(F)$ converges and its limit is equal to the Haar measure λ. □

5.2. Regeneration of measures point of view

In this subsection we present the probabilistic approach proposed by P. Ferrari, A. Maass, S. Martínez and P.Ney in [6]. Further works that use the same idea are [10, 16].
Let μ be any shift invariant probability measure on a fullshift $A^{\mathbb{Z}}$ and consider $w = (\dots, w_{-2}, w_{-1}) \in A^{-\mathbb{N}}$ (for our purposes $-\mathbb{N} = \{\dots -4, -3, -2, -1\}$). We denote by μ_w the conditional probability measure on $A^{\mathbb{N}}$.

Definition 5.3. One says that μ has *complete connections* if given $a \in A$ and $w \in A^{-\mathbb{N}}$, $\mu_w([a]_0) > 0$. If μ is a probability measure with complete connections, one defines for every $m \geq 1$

$$\gamma_m = \sup \left(\left| \frac{\mu_v([a]_0)}{\mu_w([a]_0)} - 1 \right| \; : \; v, w \in A^{-\mathbb{N}};\; v_{-i} = w_{-i},\; 1 \leq i \leq m \right) .$$

In addition, if $\sum_{m \geq 1} \gamma_m < \infty$ one says μ has *summable decay of correlations*.

The main result in [6] states:

Theorem 5.2 (Ferrari, Maass, Martínez, Ney)**.** *Let $(A, +)$ be a finite Abelian group with $|A| = p^s$ with p prime. Let $F : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be the CA given by $id + \sigma$. If μ is a probability measure on $A^{\mathbb{Z}}$ with complete connections and summable decay of correlations, then for all $w \in A^{-\mathbb{N}}$ it holds*

$$\mathcal{M}_\mu(F) = \lim_{N \to \infty} \mathcal{M}_\mu^N(F) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} F^n \mu = \lambda \, .$$

We observe that the technical conditions on the cardinality of A and the form of F can be dropped to state a result similar to that in Theorem 5.1. Historically this is the first large generalization (in relation to the class of initial measures) to Lind's result.

5.2.1. Main Idea behind the proof: regeneration of measures

Beyond the study of the combinatorial properties of the Pascal triangle modulo a prime number the main ingredient in the proof of Theorem 5.2 is the representation of a probability measure with complete connections and summable decay of correlations by means of uniform independent variables. The idea behind is trying to *mimic* the computations made in the Bernoulli case.

Let $(T_i : i \geq 1)$ be an increasing sequence of non-negative integer random variables. For every finite subset L of $\mathbb{N}$ define

$$\mathbf{N}(L) = \left| \{ i \geq 1 : T_i \in L \} \right| .$$

One says that $(T_i : i \geq 1)$ is a *stationary renewal process with finite mean interrenewal time* if

(1) $(T_i - T_{i-1} : i \geq 2)$ are independent identically distributed with finite expectation, they are independent of T_1 and $\mathbb{P}(T_2 - T_1 > 0) > 0$;

(2) For $n \in \mathbb{N}$, $\mathbb{P}(T_1 = n) = \frac{1}{\mathbb{E}(T_2 - T_1)}\mathbb{P}(T_2 - T_1 > n)$.

These conditions imply the stationary property: for every finite subset L of $\mathbb{N}$ and every $a \in \mathbb{N}$ the random variables $\mathbf{N}(L)$ and $\mathbf{N}(L + a)$ have the same distribution.

Theorem 5.3 (Ferrari, Maass, Martínez, Ney, [6])**.** *Let μ be a shift invariant probability measure on $A^{\mathbb{Z}}$ with complete connections and summable decay of correlations. There exists a stationary renewal process $(T_i : i \geq 1)$ with finite mean interrenewal time such that for every $w \in A^{-\mathbb{N}}$, there exists a random sequence $z = (z_i : i \geq 1)$ with values in A and distribution μ_w such that $(z_{T_i} : i \geq 1)$ are i.i.d. uniformly distributed in A and independent of $(z_i : i \in \mathbb{N} \setminus \{T_1, T_2, \ldots\})$.*

From the construction of the renewal process in [6] one also gets the following properties:

(1) There exists a function $\rho : \mathbb{N} \to \mathbb{R}$ decreasing to zero such that $\mathbb{P}(\mathbf{N}(L) = 0) \leq \rho(|L|)$, for any finite subset L of $\mathbb{N}$.

(2) Given $n, \ell \in \mathbb{N} \setminus \{0\}$, $1 \leq k_1 < ... < k_\ell \leq n$ and $j_1, ..., j_l \in \mathbb{N}$, for all $a_1, ..., a_n \in A$,

$$\mu_w\left(z_i = a_i, i \in \{1, ..., n\}; T_{j_1} = k_1, ..., T_{j_\ell} = k_\ell\right) =$$

$$\frac{1}{|A|^\ell}\mu_w\left(z_i = a_i, i \in \{1, ..., n\} \setminus \{k_1, ..., k_\ell\}; T_{j_1} = k_1, ..., T_{j_\ell} = k_\ell\right) .$$

(3) For any $n \in \mathbb{N}$ and $v \in A^*$, $\mu_w\left(\{\mathbf{N}(\{0, \ldots, n-1\}) > 0\} \cap [v]_n\right)$ does not depend on $w \in A^{-\mathbb{N}}$.

In the language of [20] these properties allow to prove that shift invariant probability measures with complete connections and summable decay of correlations on $A^{\mathbb{Z}}$ are harmonically mixing, and thus one can conclude Theorem 5.2. The proof in [6] did not follow this path explicitly, nevertheless they are analogous.

Theorem 5.4 (Host, Maass, Martínez, [10]; use ideas in [6])**.** *A shift invariant probability measure with complete connections and summable decay of correlations on $A^{\mathbb{Z}}$ is harmonically mixing.*

Sketch of the proof. Let μ be a shift invariant probability measure on $A^{\mathbb{Z}}$ with complete connections and summable decay of correlations. Fix a past sequence $w \in A^{-\mathbb{N}}$ and let $(T_i : i \geq 1)$ be the renewal process induced by μ. We write $\mathbb{P}_w$ for the probability, when the random variables $(z_i : i \in \mathbb{N})$ in Theorem 5.3 are given the distribution μ_w. The probability measure $\mathbb{P}$ is the integral of $\mathbb{P}_w$ with respect to $w \in A^{-\mathbb{N}}$.
For a finite subset R of $\mathbb{Z}$ and $\mathbf{x} \in A^{\mathbb{Z}}$ we write x_R for the sequence $(x_i : i \in R)$ in A^R.
Let $\chi : A^{\mathbb{Z}} \to \mathbb{T}^1$ be a character. There exist a finite set $R \subset \mathbb{Z}$ and a sequence $(\chi_n : n \in \mathbb{Z})$ in $\widehat{A}$, with $\chi_n = 1$ for $n \notin R$, $\chi_n \neq 1$ for $n \in R$ and $\chi(x) = \prod_{n \in \mathbb{Z}} \chi_n(x_n)$ for every $\mathbf{x} \in A^{\mathbb{Z}}$. We have to find an upper bound for $|\int \chi \, d\mu|$ depending only on $|R|$. As μ is shift invariant we can assume that $R \subset \mathbb{N}$.

For any finite subset R' of $\mathbb{Z}$ define $\chi_{R'} : A^{R'} \to \mathbb{T}^1$ by $\chi_{R'}(y) = \prod_{r\in R'} \chi_r(y_r)$. Observe that $\chi_{R'}(A^{R'})$ is a subgroup of $\mathbb{T}^1$ and that we can identify χ with χ_R because $\chi_R(x_R) = \chi(\mathbf{x})$ for $\mathbf{x} \in A^{\mathbb{Z}}$.
Denote $\Xi = \chi_R(A^R)$ and define $\tau(R) = \inf\{i \in R : \mathbf{N}(\{i\}) = 1\}$, where $\inf \emptyset = \infty$. We have

$$\int_{A^{\mathbb{Z}}} \chi(x_R) d\mu_w(\mathbf{x}) = \sum_{\xi\in\Xi} \xi\, \mu_w(\chi(x_R) = \xi)$$
$$= \sum_{i\in R} \sum_{\xi\in\Xi} \xi\, \mathbb{P}_w(\chi(z_R) = \xi, \tau(R) = i) + \sum_{\xi\in\Xi} \xi\, \mathbb{P}_w\big(\chi(z_R) = \xi, \tau(R) = \infty\big)\,.$$

Let $i \in R$ and set $R_i = R \setminus \{i\}$, so $\Xi = \chi_{R_i}(A^{R_i})\chi_i(A)$. For $\xi \in \Xi$ we define $V_i(\xi) = \chi_{R_i}^{-1}(\xi\chi_i(A))$. A word $y = (y_r : r \in R_i) \in A^{R_i}$ belongs to $V_i(\xi)$ if and only if there exists $a \in A$ such that the word y' obtained from y by putting $y'_i = a$ satisfies $\chi(y') = \xi$.
For $y \in A^{R_i}$ we put $\xi_y = \chi_{R_i}(y)$ and for $\xi \in \Xi$ we define $V_i(\xi, y) = \chi_i^{-1}(\xi\xi_y^{-1})$. Since $V_i(\xi, y)$ is a coset of $Ker(\chi_i)$, we get $|V_i(\xi, y)| = |Ker(\chi_i)|$. Therefore,

$$\sum_{\xi\in\Xi} \xi\, \mathbb{P}_w\big(\chi(z_R) = \xi, \tau(R) = i\big)$$
$$= \sum_{\xi\in\Xi} \xi \sum_{y\in V_i(\xi)} \sum_{y_i\in V_i(\xi,y)} \mathbb{P}_w\big(z_r = y_r,\, r \in R; \tau(R) = i\big)$$
$$= \sum_{\xi\in\Xi} \xi \sum_{y\in V_i(\xi)} \sum_{y_i\in V_i(\xi,y)} \frac{1}{|A|}\, \mathbb{P}_w\big(z_r = y_r,\ r \in R_i;\ \tau(R) = i\big)$$
$$= \sum_{\xi\in\Xi} \xi \sum_{y\in V_i(\xi)} \frac{|V_i(\xi, y)|}{|A|}\, \mathbb{P}_w\big(z_r = y_r,\ r \in R_i;\ \tau(R) = i\big)$$
$$= \frac{|Ker(\chi_i)|}{|A|} \sum_{\xi\in\Xi} \xi \sum_{y\in V_i(\xi)} \mathbb{P}_w\big(z_r = y_r,\ r \in R_i;\ \tau(R) = i\big)$$

$$= \frac{|Ker(\chi_i)|}{|A|} \sum_{y\in A^{R_i}} \mathbb{P}_w\big(z_r = y_r,\ r \in R_i;\ \tau(R) = i\big) \cdot \sum_{\{\xi\in\Xi : V_i(\xi,y)\neq\emptyset\}} \xi$$

where in the second equality we have used Theorem 5.3. Recall $\xi_y = \chi_{R_i}(y)$. We have

$$\{\xi \in \Xi : V_i(\xi, y) \neq \emptyset\} = \{\xi \in \Xi : \chi_i^{-1}(\xi\xi_y^{-1}) \neq \emptyset\} = \xi_y\chi_i(A)\,.$$

Hence

$$\sum_{\{\xi\in\Xi : V_i(\xi,y)\neq\emptyset\}} \xi = \xi_y \sum_{\xi\in\chi_i(K)} \xi = 0\,.$$

We conclude

$$\sum_{\xi\in\Xi} \xi\ \mathbb{P}_w\big(\chi(r_R)=\xi,\ \tau(R)=i\big)=0\ .$$

Coming back to the integral we get,

$$\begin{aligned}\Big|\int_{A^{\mathbb{Z}}}\chi(x_R)d\mu_w(\mathbf{x})\Big| &= \Big|\sum_{\xi\in\Xi}\xi\ \mathbb{P}_w\big(\chi(x_R)=\xi,\ \tau(R)=\infty\big)\Big| \\ &\leq |\Xi|\,\mathbb{P}_w\big(\tau(R)=\infty\big)\leq |K|\,\rho(|R|)\ .\end{aligned}$$

Since this inequality holds for any $w\in A^{-\mathbb{N}}$ we have $\big|\int\chi(\mathbf{x})\,d\mu(\mathbf{x})\big|\leq |A|\,\rho(|R|)$. Since $\rho(|R|)\to 0$ as $|R|\to\infty$ we conclude that μ is harmonically mixing. □

5.3. Some generalizations

There are several extensions of Theorem 5.3. Just to give the flavour of them here we give one where fullshifts are changed by subgroup shifts. Other generalizations appeared in [16, 24].

Theorem 5.5 (Maass, Martínez, Pivato, Yassawi, [14, 15])**.** *Let $\mathcal{G}\subseteq A^{\mathbb{Z}}$ be an irreducible subgroup shift verifying the following-lifting-property (resp. A is p^s-torsion with p prime). Let $F:\mathcal{G}\to\mathcal{G}$ be a proper linear block map and μ a probability measure with complete connections and summable memory decay compatible with $\mathcal{G}$. Then, the Cesàro mean of μ under the action of F converges to the Haar measure of $\mathcal{G}$. If A is a p-group with p-prime $\mathcal{G}$ always verifies the FLP property.*

§6. Final comments and questions

– We would like to change the "complete connections and summable decay of correlations" property by some *mixing* property for the shift map.

– The asymptotic randomization does not require full support of the initial measure and positive entropy w.r.t. the shift map: there exist shift invariant measures μ on $\{0,1\}^{\mathbb{Z}}$ with $h_\mu(\sigma)=0$ that are asymptotically randomized by $F=id+\sigma$ (see Pivato and Yassawi examples in [22]).

— *Question:* Do Cesàro means exist for expansive and/or positively expansive block-maps of a mixing shift of finite type ? how the limit is related with the unique maximal entropy measure ? There are some partial results for classes of right permutative cellular automata: with associative local rules, or N-scaling local rules (see [10]). These CA can be seen as the product of an algebraic CA with a shift. Here measures are not asymptotically randomized but the limits are the product of a maximal measure with a periodic measure, so combining results from [6, 20, 21] and [2].

Acknowledgements

The author was supported by CMM-Fondap-Basal fund.

References

[1] F. BLANCHARD, A. MAASS. Dynamical properties of expansive one-sided cellular automata. *Israel J. Math.* **99** (1997), 149–174.

[2] F. BLANCHARD, P. TISSEUR. Some properties of cellular automata with equicontinuity points. *Ann. Inst. H. PoincarŐ Probab. Statist.* **36** (2000), no. 5, 569–582.

[3] M. BOYLE, A. MAASS. Expansive invertible onesided cellular automata. *J. Math. Soc. Japan* **52** (2000), no. 4, 725–740.

[4] E. COVEN, M. PAUL. Endomorphisms of irreducible subshifts of finite type. *Math. Systems Theory* **8** (1974/75), no. 2, 167–175.

[5] M.EINSIEDLER. Isomorphism and measure rigidity for algebraic actions on zero-dimensional groups. *Monatsh. Math.* **144** (2005), no. 1, 39–69.

[6] P. FERRARI, A. MAASS, S. MARTÍNEZ, P. NEY. Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems* **20** (2000), no. 6, 1657–1670.

[7] E. GLASNER, B. WEISS. Sensitive dependence on initial conditions. *Nonlinearity* **6** (1993), no. 6, 1067–1075.

[8] A. HEDLUND. Endormorphisms and automorphisms of the shift dynamical system. *Math. Systems Theory* **3** (1969), 320–375.

[9] B. HOST. Nombres normaux, entropie, translations. *Israel J. Math.* **91** (1995), no. 1-3, 419–428.

[10] B. HOST, A. MAASS, S. MARTÍNEZ. Uniform Bernoulli measure in dynamics of permutative cellular automata with algebraic local rules. *Discrete Contin. Dyn. Syst.* **9** (2003), no. 6, 1423–1446.

[11] D. LIND, B. MARCUS. An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge, 1995. xvi+495 pp.

[12] D. LIND. Applications of ergodic theory and sofic systems to cellular automata. Cellular automata (Los Alamos, N.M., 1983). *Phys. D* **10** (1984), no. 1-2, 36–44.

[13] A. MAASS, S. MARTÍNEZ. On Cesàro limit distribution of a class of permutative cellular automata. *J. Statist. Phys.* **90** (1998), no. 1-2, 435–452.

[14] A. MAASS, S. MARTÍNEZ, M. PIVATO, R. YASSAWI. Asymptotic randomization of subgroup shifts by linear cellular automata. *Ergodic Theory Dynam. Systems* **26** (2006), no. 4, 1203–1224.

[15] A. MAASS, S. MARTÍNEZ, M. PIVATO, R. YASSAWI. Attractiveness of the Haar measure for linear cellular automata on Markov subgroups. Dynamics & stochastics, 100–108, IMS Lecture Notes Monogr. Ser., 48, Inst. Math. Statist., Beachwood, OH, 2006.

[16] A. MAASS, S. MARTÍNEZ, M. SOBOTTKA. Limit measures for affine cellular automata on topological Markov subgroups. *Nonlinearity* **19** (2006), no. 9, 2137–2147.

[17] M. NASU. Textile systems for endomorphisms and automorphisms of the shift. *Mem. Amer. Math. Soc.* **114** (1995), no. 546, viii+215 pp.

[18] D. RUDOLPH. $\times 2$ and $\times 3$ invariant measures and entropy. *Ergodic Theory Dynam. Systems* **10** (1990), no. 2, 395–406.

[19] K. PETERSEN. Ergodic theory. *Cambridge Studies in Advanced Mathematics* **2**. Cambridge University Press, Cambridge, 1983. xii+329 pp.

[20] M. PIVATO, R. YASSAWI. Limit measures for affine cellular automata. *Ergodic Theory Dynam. Systems* **22** (2002), no. 4, 1269–1287.

[21] M. PIVATO, R. YASSAWI. Limit measures for affine cellular automata. II. *Ergodic Theory Dynam. Systems* **24** (2004), no. 6, 1961–1980.

[22] M. PIVATO, R. YASSAWI. Asymptotic randomization of sofic shifts by linear cellular automata. *Ergodic Theory Dynam. Systems* **26** (2006), no. 4, 1177–1201.

[23] M. PIVATO. Invariant measures for bipermutative cellular automata. *Discrete Contin. Dyn. Syst.* **12** (2005), no. 4, 723–736.

[24] M. PIVATO. Module shifts and measure rigidity in linear cellular automata. *Ergodic Theory Dynam. Systems* **28** (2008), no. 6, 1945–1958.

[25] M. SABLIK. Measure rigidity for algebraic bipermutative cellular automata. *Ergodic Theory Dynam. Systems* **27** (2007), no. 6, 1965–1990.

[26] S. SILBERGER. Subshifts of the three dot system. *Ergodic Theory Dynam. Systems* **25** (2005), no. 5, 1673–1687.

Departamento de Ingeniería Matemática,
Universidad de Chile & Centro de Modelamiento Matemático
UMI 2071 UCHILE-CNRS
Casilla 170/3 correo 3
Santiago, Chili.
amaass@dim.uchile.cl

Contributed Papers

Monografías Matemáticas García de Galdeano **36**, 59–67 (2010)

ON THE ADMISSION-ABANDON CONTROL OF THE M/M/S/K+G QUEUE

F. Avram, D. F. Chedom and L. P. Fotso

Abstract. Motivated by the application to call centers, we consider a multiserver queue with waiting room of finite size K, where customers who find all the servers busy may leave before the beginning of their service, according to a general impatience distribution $G(\cdot)$ of mean D. For this model, denoted by M/M/s/K + G, we investigate numerically the optimal choice of the parameters K, D, focusing on the case of deterministic impatience D, for several objectives involving reneging, balking, waiting and refusal costs.

Keywords: M/M/s/K + G queue, impatience, Erlang C, finite buffer, admission control.

AMS classification: 60K25, 93E20, 68M20

§1. Introduction

The model. Consider the $M/M/s/K/+G$ model, where arrivals occur according to a Poisson process with rate λ and service times are exponentially distributed with mean μ^{-1}. The number of servers is s, the buffer (number of waiting places) limit is K and an arrival customer who finds $s+K$ customers in the system is rejected. Customers are served in the order of arrival, if they don't abandon. Furthermore, customers arrive with deadlines until the beginning of service, which are i.i.d. random variables $I_1, I_2, ...$, with a general "patience distribution" $G(\cdot)$ with mean D. Balking (refusal to enter if not served immediately) is modeled by introducing an atom of mass $p = P[I_k = 0]$ in the patience distribution. Furthermore, customers may abandon the system (or renege) after entering, if service does not start before their deadline.

Markovian modelisation. Let $N(t)$ be the total number of customers in the system at time t. It is convenient to consider separately three cases: i) Unsaturated regime: $N(t) < s$, ii) Saturated regime $N(t) = s$, iii) Overloaded regime $N(t) \geq s+1$.

Let $U(t)$ denote the time until the waiting buffer at time t becomes unsaturated, also called virtual offered waiting time (v.o.w.t). Thus, a customer who arrives at time t with infinite patience (i.e. no deadline) must wait $U(t)$ before his service. The distribution of $U(t)$ conditional on the number $K(t) = (N(t) - s + 1)_+ = k$ is Erlang with k stages and scale parameter $\mu_s = s\mu$.

Aware customers. There is a model with "aware customers", who know the v.o.w.t. at their arrival, and use this to leave immediately, if it exceeds their deadline. We may specify the "aware" Markovian system by:

$$Z(t) := (U(t), N(t)) = \begin{cases} (0, i) & \text{if } N(t) = i, 0 \leq i < s, \\ (U(t), i) & \text{if } N(t) = i, i \geq s. \end{cases} \tag{1}$$

Notes: The saturated regime, starting when all servers become busy and ending with the first arrival which must wait, may be viewed as a common boundary state, from which the process may be "teleported" to either the unsaturated states $N(t) < s$, or to the overloaded regime $N(t) \geq s+1$, with rates $s\mu$ and λ, respectively.

Unaware customers. There is also an alternative model where the i-th customer admitted at time t will leave at time $t + I_i$ if $I_i < U(t)$. This model leads to a worse utilization of the waiting buffer and to a bigger admittance refusal probability, but the offered waiting time and abandon probability are the same [2], since for a customer already admitted in the line, the impatient leaving customers might as well have left at their arrival (and so don't affect $U(t)$).

Historical notes. The pioneer work on impatience is Palm [10], who proposed both the exponential and the Weibull as reasonable distributions for patience times. Deterministic impatience has first been studied by Barrer [3] and Gnedenko and Kovalenko [8], who employed a Markovian model specifying the joint behavior of the elapsed waiting time in line $X(t)$ of the queue's first customer, and the total number of customers waiting $N(t)$. When $K = \infty$, $X(t)$ alone provides a Markovian description of the system in waiting regime (and the number of waiting customers $(N(t) - s)_+$ is Poisson distributed), as noticed by Choi & al [7].

The case of general impatience was studied by Baccelli, Boyer and Hebuterne – see for example [2], who used the alternative modelization via $U(t)$.

The combined effect of impatience and refusal of customers was considered in Ancker and Gafarian [1] and Subba Rao [11], who assumed exponential impatience. Most recently, extensive analyses of the model with state dependent rates and general impatience have been provided by Movaghar [9] and Brandt and Brandt [4]. See also Zeltyn and Mandelbaum [12] for further references and developments.

The contribution of our work consists in studying numerically the choice of the parameters K, D for optimal admission control under this model, for several objectives involving abandonment, waiting and refusal costs. The rest of the paper is structured as follows. In section 2 we review, following the results of Baccelli & al [2], Movaghar [9], and Brandt [4], first the characteristics of the classical model with $K = \infty$, and then, in subsection 2.3, for the case $K < \infty$. The special case of deterministic impatience is presented in section 3, where we introduce also the performance measures to be optimized. Finally, we deal with optimization in section 4.

§2. Performance characteristics

2.1. The undersaturated and saturated regimes

The queue evolves in this "immediate service" regime as a classical Markovian birth and death process. The stationary distribution: $q_i = P\{N(t) = i\}, i = 0, ..., s$, satisfies the standard equilibrium equations:

$$(\lambda + i\mu)q_i = \lambda q_{i-1} + (i+1)\mu q_{i+1}, \;\; 0 \leq i \leq s-1, \tag{2}$$

(where $q_{-1} = 0$) which reduce to local equilibrium equation $i\mu q_i = \lambda q_{i-1}, \;\; 0 \leq i \leq s$. Putting $\rho = \lambda/\mu$, we find:

$$q_i = q_0 \frac{\rho^i}{i!}, \quad 0 \leq i \leq s. \tag{3}$$

2.2. The waiting regime

In this regime, the basic quantity which determines the various performance metrics of the queue is the stationary density of the v.o.w.t $U(t)$:

$$v(x) = \frac{d}{dx} \left(P[N(t) \geq s, U(t) \leq x] \right), \quad x \in \mathbb{R}^+. \tag{4}$$

It may then be checked by the method of level crossings [5], [6] (see also (3.30) in [9]) that $v(x)$ satisfies the integral equation:

$$v(x) = \lambda q_{s-1} e^{-\mu_s x} + \lambda \int_0^x \bar{G}(y)\, v(y) e^{-\mu_s (x-y)} dy, \tag{5}$$

with solution:

$$v(x) = v(0) e^{\lambda\, H(x) - \mu_s x}, \qquad v(0) = \lambda\, q_{s-1} = \mu_s q_s, \tag{6}$$

where $H(x) = E[min(x, I)] = \int_0^x \bar{G}(y) dy$ (where I is a customer's impatience time) represent the expected time a given customer spent waiting during $(0, x]$.

Remark 1. Note that $H(\infty) = D$, the mean patience time.

Remark 2. The stationary probabilities of a client abandoning and being served are respectively: $P_{Ab} = \int_0^\infty v(x) G(x) dx, \quad p_{Ser} = \int_0^\infty v(x) \bar{G}(x) dx$.

Using (6), we can compute now q_0 from the normalization condition in the case $K = \infty$, completing thus the description of the stationary measure:

$$\sum_{i=0}^{s-1} q_i + \int_0^\infty v(x) dx = q_0 \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \lambda \frac{\rho^{s-1}}{(s-1)!} \int_0^\infty e^{\lambda H(x) - \mu_s x} dx \right) = 1. \tag{7}$$

Putting now $J = \displaystyle\int_0^\infty \mu_s e^{\lambda\, H(x) - \mu_s x} dx = E e^{\lambda \min[I, S]}$, where S denotes the time until the first departure from service, we find (compare to (3.3) in [4] and (3.27) in [9]) that :

$$q_0 = \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \frac{\rho^{s-1}}{(s-1)!}\, \psi\, J \right)^{-1} = \frac{1}{\frac{\rho^{s-1}}{(s-1)!} (\mathcal{E}_s + \psi J)} \tag{8}$$

where $\psi = \lambda/\mu_s$, and where $\mathcal{E}_s = \frac{\sum_{i=0}^{s-1} \frac{\rho^i}{i!}}{\frac{\rho^{s-1}}{(s-1)!}} = \int_0^\infty e^{-t} (1 + t/\rho)^{s-1} dt := h_\Gamma^{-1}(s, \rho)$ is the inverse of the Gamma distribution hazard function (which yields in queueing theory the fundamental Erlang-B formula).

2.3. The finite buffer model

When $K < \infty$, we must also keep track of the buffer occupation, and consider therefore the joint stationary density of the v.o.w.t. $U(t)$ and the number k of customers waiting:

$$v(k,x) = \frac{d}{dx}\left(P[N(t) - s = k, U(t) \leq x]\right) \tag{9}$$
$$= q_s \mu_s e^{-\mu_s x} \frac{(\lambda H(x))^k}{k!} = v_\infty(x) \frac{e^{-\lambda H(x)} (\lambda H(x))^k}{k!}, \quad k = 0, 1, 2, ...K$$

This basic formula does not seem to appear explicitly in the literature, but is equivalent of course to the known formulas for the conditional density $v(x|k)$ and marginal $p_k = \int_0^\infty v(k,x)dx$ recalled in the notes below.

Notes: 1) From formula (9), we obtain immediately the marginal stationary distributions of $U(t)$ and of the number k of customers waiting. Letting

$$\Phi_k = \int_0^\infty H(x)^k \mu_s e^{-\mu_s x} dx = EH[S]^k, \quad k = 0, 2, ..., K \tag{10}$$

we find that the stationary distribution of the number k of customers waiting is:

$$p_k = P[N(t) = s + k] = \int_0^\infty v(k,x)dx = q_s \frac{\lambda^k}{k!} \Phi_k, \quad k = 0, 1, 2, ...K, \tag{11}$$

which coincides with (3.1) of [4] and with (3.33) of [9] (since $\lambda q_{s-1} = \mu_s q_s$, and $p_0 = q_s$).
2) The density of the v.o.w.t. $U(t)$ conditional on the number k of customers waiting is:

$$v(x|k) = \frac{v(k,x)}{p_k} = \mu_s e^{-\mu_s x} \frac{H(x)^k}{\Phi_k}, \quad k = 0, 1, 2, ...K. \tag{12}$$

which is the same result as (3.17) in [9] [§] .
3) Brandt and Brandt[4] and Movaghar[9] showed furthermore that the intensity of customers leaving due to impatience conditional on the number k of customers waiting is given by: $\theta_k = \frac{k\Phi_{k-1}}{\Phi_k} - \mu_s$.
4) From the normalization condition:

$$\sum_{i=0}^{s-1} q_i + \sum_{k=0}^{K} p_k = q_0 \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \frac{\rho^s}{s!} \sum_{k=0}^{K} \frac{\lambda^k}{k!} \Phi_k \right) = 1, \tag{13}$$

we obtain q_0:

$$q_0 = \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \frac{\rho^s}{s!} \sum_{k=0}^{K} \frac{\lambda^k}{k!} \Phi_k \right)^{-1} = \left(\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \frac{\rho^{s-1}}{(s-1)!} \psi J_K \right)^{-1} \tag{14}$$

where $J_K = \sum_{k=0}^{K} \frac{\lambda^k}{k!} \Phi_k$, which coincides with (3.35) in [9].

[§]The formula (9) states that under the stationary measure, the variable $U(t)$ and the number of arrivals during $H(U(t))$ are independent, and thus $v(k,x)$ is the product of the probability $v(x)$ and of the probability of having k arrivals during $H(x)$, which is $e^{-\lambda H(x)} (\lambda H(x))^k / k!$

§3. The special case with deterministic impatience: the M/M/s/K + D queue

From the above results, we have the following ones for queues with deterministic impatience D :

$$H(x) = min[x, D] = \begin{cases} x, & 0 \le x \le D, \\ D, & x > D, \end{cases}, \qquad \Phi_k = \frac{\tilde{\Gamma}_k(\mu_s D)}{\mu_s^k}, \tag{15}$$

where $\tilde{\Gamma}_k(x) = 1 - e^{-x} \sum_{j=0}^{k-1} \frac{(x)^j}{j!} = 1 - \frac{\Gamma(k,x)}{\Gamma(k)}$, and where $\Gamma(k)$ and $\Gamma(k,x)$ represent the Gamma and incomplete Gamma function (indeed, the integral (10) is easily checked ...)
It follows that the joint density of the v.o.w.t. $U(t)$ and of the number k of customers waiting is

$$v(k,x) = \begin{cases} q_s \mu_s e^{-\mu_s x} \frac{(\lambda x)^k}{k!}, & x \le D, \\ q_s \mu_s e^{-\mu_s x} \frac{(\lambda D)^k}{k!}, & x > D, \end{cases} \tag{16}$$

In the unsaturated regime, we have: $P\{N(t) = n\} = q_n, \quad 0 \le n \le s-1$, and in the saturated regime $P\{N(t) = n\} = \int_0^\infty v(n-s, x)dx = q_s \frac{\psi^{n-s}}{(n-s)!} \tilde{\Gamma}_{n-s}(\mu_s D)$.
Letting now $P_n = P\{N(t) = n\}, \quad 0 \le n \le s+K$ be the system size distribution, we recover (4.5) of [9]:

$$P_n = \begin{cases} q_n, & N(t) = n, \quad 0 \le n \le s-1, \\ q_s \frac{\psi^{n-s}}{(n-s)!} \tilde{\Gamma}_{n-s}(\mu_s D), & N(t) = n, \quad s \le n \le s+K. \end{cases} \tag{17}$$

with $q_0 = \left(\sum_{i=0}^{s} \frac{\rho^i}{i!} + \frac{\rho^s}{s!} \sum_{k=1}^{K} \lambda^k \Phi_k \right)^{-1}$.
The stationary distribution of the number k of customers waiting is:

$$p_k = P_{s+k} = q_s \psi^k \tilde{\Gamma}_k(\mu_s D), \quad k = 0, 1, 2, ...K; \tag{18}$$

Finally, the density of the v.o.w.t. $U(t)$ conditional on the number k of customers waiting is:

$$v(x|k) = \frac{v(k,x)}{p_k} = \begin{cases} \frac{\mu_s e^{-\mu_s x} (\lambda x)^k}{k! \psi^k \tilde{\Gamma}_k(\mu_s D)}, & x \le D, \\ \frac{\mu_s e^{-\mu_s x} (\lambda D)^k}{k! \psi^k \tilde{\Gamma}_k(\mu_s D)}, & x > D, \end{cases} \tag{19}$$

which is the same result as (4.4) in [9].
Some performance measures with deterministic impatience are:

1. The probability of being refused entry is :

$$P_{Ref} = P_{s+K} = q_s \psi^K \tilde{\Gamma}_K(\mu_s D); \tag{20}$$

2. The stationary loss probability

$$P_{Ab}(D) = \frac{\sum_{k=1}^{K} v(k,D)}{\lambda} = q_s e^{-\mu_s D} \sum_{k=0}^{K-1} \frac{(\lambda D)^k}{k!} = q_s e^{(\lambda-\mu_s) D} \frac{\Gamma(K, \lambda D)}{\Gamma(K)}. \quad (21)$$

Note that when $K = \infty$ and $D = 0$ we recover the Erlang-B formula $\frac{\rho^s/s!}{\sum_{i=0}^{s} \rho^i/i!}$;

3. The mean queue size E_Q is given by :

$$E_Q \quad = \quad \sum_{k=1}^{K} k\, P_{s+k} = \sum_{k=1}^{K} k\, q_s\, \psi^k \tilde{\Gamma}_k(\mu_s D). \quad (22)$$

After some tedious algebra we obtain:

$$E_Q = q_s\, \psi \left(\frac{1 - e^{-\mu_s D} + g(K, \mu_s D)\, \psi^K((K+1)\psi - K - 1 - \psi)}{(1-\psi)^2} \right), \quad (23)$$

where $g(k,x) = 1 - \dfrac{\Gamma(k+1,x)}{k!}$.

Taking $D \to \infty$ we get the expression of the standard M/M/s/K model:

$$E_Q = q_s\, \psi \left(\frac{1 + K\psi^{K+1} - (K+1)\psi^K}{(1-\psi)^2} \right); \quad (24)$$

4. The mean waiting time comes from the Little formula:

$$E_W = \frac{E_Q}{\lambda^*} = q_s\, \psi \left(\frac{1 + K\psi^{K+1} - (K+1)\psi^K}{\lambda^*\, (1-\psi)^2} \right), \quad (25)$$

where $\lambda^* = \lambda(1 - P_{Ref})$.

§4. Optimization

Let w denote the waiting cost incured each time a customer spend in the queue, c_a, the abandoning cost incurred each time a waiting customer abandons the system before his service has begun and let c_r denote the rejection cost incurred each time an arriving customer is rejected at entry. Assuming already an optimal number of servers s, our optimization goal here consists of determining the optimal D and K that minimize the customer's long-run stationary inconvenience costs, which combine waiting, abandon and rejection at entry.

$$f = w\, E_Q + c_r\, \lambda\, P_{Ref} + c_a\, \lambda(1 - P_{Ref})\, P_{Ab}$$

In this preliminary study, we consider only the one dimensional optimizations obtained by fixing K and D, respectively, and only for the case of deterministic impatience, reviewed in Section 3.

4.1. Optimal D on an objective combining E_W and P_{Ab}

We assume now that K is fixed, and consider the optimization objective combining the waiting cost and the abandoning cost, given by:

$$f_1 = wE_Q + \lambda^* c_a P_{Ab} = w\lambda^*(E_W + a' P_{Ab}) \tag{26}$$

where $\lambda^* = \lambda(1 - P_{Ref}), a' = \dfrac{c_a}{w}$ and where we applied Little's formula.
Since λ^* is a positive constant when K is fixed, minimizing f_1 is equivalent to minimizing $\xi_1 = E_W + a' P_{Ab}$
The ratio a' between the abandoning cost and the waiting cost is an important management decision parameter, reflecting what is more expensive: abandoning or waiting.
Note that E_W increases with D while P_{Ab} decreases with D, as shown in figure 1, obtained with $\lambda = \mu = 1$, $s = 2$, and the queue limit size $K = 10$. While ξ_1 (and thus f_1) need not be convex in D in general, we found this to be the case often, like for example with $\lambda = \mu = 1$, $s = 2$, and $K = 10$, and for any $a' \in [1.0, 1.5]$. When $a' = 1.2$ our combined objective admits a minimum $D^* = 0.56$ as shown in figure 2

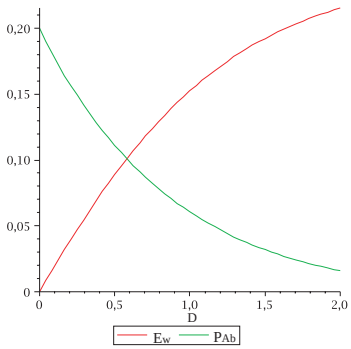


Figure 1 : E_W and P_{Ab} depending on D, for $\lambda = 1, s = 2, K = 10$

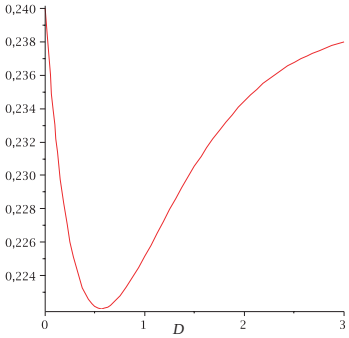


Figure 2: The combined objective of E_w and P_{Ab} depending on D, for $\lambda = 1, s = 2$, $K = 10$, and $a' = 1.2$

Table (1) gives different values of D^* for $\mu = 1$ and a set of given couples (λ, a') with $s = 2$ and $K = 10$ [§] . A zero-value or infinite value means that ξ_1 does not have an interior minimum in the case considered.
Table (1) also shows that D^* increases with a', that is as the abandoning cost prevails over the waiting cost.

4.2. Optimal K on an objective combining P_{Ab} and P_{Ref}

The second optimization objective function combines the abandoning cost and the balking cost The combined objective is given by :

$$f_2 = c_a\ \lambda^*\ P_{Ab} + c_r\ \lambda\ P_{Ref} = c_a \lambda \xi_2 \tag{27}$$

[§] As $\mu = 1$, λ represents also the traffic intensity ρ

	a'								
	0.6	0.8	1	1.1	1.2	1.3	1.4	1.5	1.6
$\lambda = 0.5$	0	0.48	1.32	1.93	2.81	4.34	∞	∞	∞
$\lambda = 0.75$	0	0.16	0.53	0.82	1.2	1.8	3.05	∞	∞
$\lambda = 1$	0	0	0.21	0.37	0.56	0.83	1.29	2.83	∞
$\lambda = 1.25$	0	0	0	0.15	0.26	0.41	0.61	1.0	∞
$\lambda = 1.5$	0	0	0	0	0.11	0.2	0.3	0.4	∞
$\lambda = 1.75$	0	0	0	0	0	0.08	0.15	0.23	∞
$\lambda = 2$	0	0	0	0	0	0	0.06	0.11	∞

Table 1: D^* for a set of given couples (λ, a') with $s = 2$ and $K = 10$.

where $\xi_2 = (1 - P_{ref})P_{Ab} + r' \; P_{ref}$, with $r' = \dfrac{c_r}{c_a}$. Of course, minimizing f_2 is equivalent to minimizing ξ_2.
The ratio r' between the rejection cost and the desertion cost is another important management decision parameter, reflecting what is more expensive: rejection or desertion.
Our goal here is to obtain the optimal K which minimizes the objective function, when D is fixed. Note that P_{Ab} increases with K while P_{Ref} decreases with K, as shown in figure 3, obtained with $\lambda = 1$, $s = 2$ and $D = 1$. When $\lambda = 1$, $s = 2$, and $D = 1$, ξ_2 (thus, f_2) is convex and admits a minimum for any $r' \in [0.1, 1.0]$. For example, with $r' = 0.5$ our combined objective admits a minimum $K^* = 1.77$ as shown in figure 4.

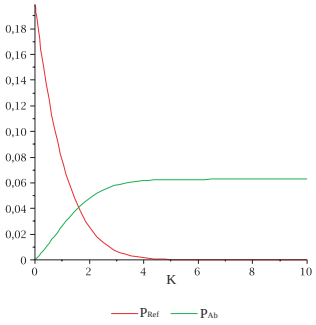


Figure 3 : P_{Ref} and P_{Ab} depending on K, for $\lambda = 1, s = 2$ and $D = 1$.

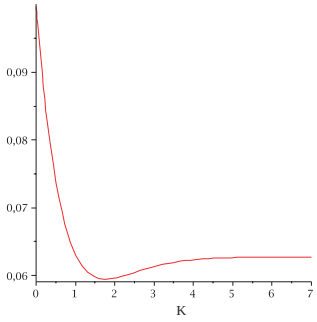


Figure 4: The combined objective of P_{Ref} and P_{Ab} depending on K, for $\lambda = 1$, $s = 2, D = 1$ and $r' = 0.5$.

Table (2) gives the different values of K^* for a set of given couples (λ, r') with $s = 2$ and $D = 1$. A zero-value or an infinite value means that ξ_2 (thus, f_2) is not convex in the case considered.
Table (2) also shows that K^* increase with r', that is when the rejection cost prevails over the waiting cost. We note that K^* doesn't strongly depend on the traffic intensity ρ.

	r'								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\lambda = 0.5$	0	0.52	0.98	1.5	2.1	3.0	4.42	∞	∞
$\lambda = 0.75$	0.05	0.47	0.88	1.35	1.93	2.71	3.92	6.21	∞
$\lambda = 1$	0.046	0.43	0.82	1.24	1.77	2.47	3.53	5.54	∞
$\lambda = 1.25$	0.051	0.43	0.81	1.22	1.71	2.37	3.36	5.22	∞
$\lambda = 1.5$	0.1	0.51	0.91	1.34	1.86	2.54	3.58	5.52	∞
$\lambda = 1.75$	0.30	0.51	1.30	1.82	2.45	3.28	4.56	$\rightarrow$	∞
$\lambda = 2$	0	0.17	0.36	0.53	0.70	0.86	1.0	1.2	1.3

Table 2: K^* for a set of given couples (λ, r') with $\mu = 1$, $s = 2$ and $D = 1$.

References

[1] C.J. ANCKER JR, AND A. GAFARIAN. Some Queueing problem with Balking and reneging. *Operations Research 11*, 1 (1963), 88–100.

[2] F. BACCELLI, P. BOYER AND G. HEBUTERNE. Single-server queues with impatient customers, *Adv. in Appl. Probab. 16* (1984), 887–905.

[3] D.Y. BARRER. Queueing with impatient customers and ordered service. *Operations Research*, 5 (1957) 650–656.

[4] A. BRANDT, M. BRANDT. On the M(n)/M(n)/s queue with impatient calls. *Performance Evaluation 35* (1999), 1–18.

[5] P.H. BRILL. An Embedded Level Crossing Technique for Dams and Queues. *Journal of Applied Probability, 16*,(1979), 174–186.

[6] P.H. BRILL AND M.J.M. POSNER. The System Point Method in Exponential Queues: A Level Crossing Approach. *Mathematics of Operations Research 6*, (1981), 31–49.

[7] B. D. CHOI, B. KIM AND J. CHUNG. M/M/1 queue with impatient customers. *Queueing Systems, 38*, (2001), 49–66.

[8] B. V. GNEDENKO AND I.N. KOVALENKO. Introduction to Queueing Theory. *Birkhauser*, (1989).

[9] A. MOVAGHAR. On Queue with customer impatience until the beginning of service. *Queueing Systems 29* (1998), 337–350.

[10] C. PALM. Methods of judging the annoyance caused by congestion. *Tele, 4* (1953), 189–208.

[11] S. SUBBA RAO. Queue with balking and reneging in M/G/1 system. *Metrika 12*, 1 (1967), 88–100.

[12] S. ZELTYN AND A .MANDELBAUM. Call centers with impatient customers: many-server asymptotics of the M/M/n+ G queue Queueing Systems. *Queueing Systems 51*, 3/4 (2005), 361–402.

Florin AVRAM
University of Pau (France)
florin.avram@univ-pau.fr

Laure Pauline FOTSO
University of Yaoundé I (Cameroon)
l_fotso@yahoo.com

Donatien CHEDOM FOTSO
University of Pau (France)
University of Yaoundé I (Cameroon)
donatien.chedomfotso@univ-pau.fr

Monografías Matemáticas García de Galdeano **36**, 69–77 (2010)

ON OPTIMAL DIVIDEND DISTRIBUTION FOR A CRAMÉR-LUNDBERG PROCESS WITH EXPONENTIAL JUMPS IN THE PRESENCE OF A LINEAR GERBER-SHIU PENALTY FUNCTION

Florin Avram, Zbigniew Palmowski and Martijn Pistorius

Abstract. In this paper we consider the optimal dividend problem in the presence of a linear Gerber-Shiu penalty function for an insurance company whose reserves evolve according to a classical risk process. We give explicit sufficient conditions for the optimality of a liquidation and barrier strategies, and analytically explicit expressions for their value functions. We use these results to completely analyze the case of exponentially distributed claims.

Keywords: Optimal control, risk process, barrier strategies, optimal dividend distribution

AMS classification: 60J99, 93E20, 60G51

§1. Introduction

In classical collective risk theory (e.g. Gerber [9]) the surplus $X = \{X_t, t \geq 0\}$ of an insurance company with initial capital x is described by the Cramér-Lundberg model:

$$X_t = x + p\,t - \sum_{k=1}^{N_t} C_k, \tag{1}$$

where C_k are independent positive random variables with common distribution function F representing the claims made, $N = \{N_t, t \geq 0\}$ is an independent Poisson process with intensity λ modeling the times at which the claims occur, and $p\,t$ represents the premium income up to time t.

The ruin problem and De Finetti's dividend problem. The classic research of the scandinavian school (Lundberg, Cramér, etc.) had focused on determining the "ruin probability" of the process (1) ever becoming negative, under the assumption that X has positive profits, i.e. that its first moment is positive. Since however in this case the surplus has the unrealistic property that it converges to infinity with probability one, De Finetti [6] introduced the dividend barrier model, in which all surpluses above a given level are transferred (subject to a discount rate) to a beneficiary, and raised the question of optimizing this barrier. An intricate "bands strategy" solution was discovered by Gerber [7], [8], as well as the fact that

for exponential claims and with no constraints on the dividends rate, this reduces to a simple barrier strategy: "pay all you can above a fixed constant barrier a". Hallin [12] formulated time dependent integro-differential equations describing the payoff associated to a $2n$ bands policy.

There has been a great deal of work on De Finetti's objective, usually concerning barrier strategies. Gerber and Shiu [10] and Jeanblanc and Shiryaev [13] consider the optimal dividend problem in a Brownian setting. Zhou [19] study the constant barrier under the Cramér-Lundberg model (1). For related work considering both excess-of-loss reinsurance and dividend distribution policies (in a diffusion setting), see Asmussen et al. [2].

Balancing dividends and ruin penalties. Since under De Finetti's objective ruin is sure and its "severity" ignored, several alternative objectives have been proposed recently, involving a continuous payoff until ruin [1], or a penalty at ruin, based on some function of the severity of ruin [18], [5], [11], [17], [15]. It is this problem that we consider below, under the Cramér-Lunderg model.

The model. The risk process before dividends are deducted is modeled by by X in (1). The risk process controlled by the dividend policy π is then given by $U^\pi = \{U_t^\pi, t \geq 0\}$, where

$$U_t^\pi = X_t - L_t^\pi, \tag{2}$$

$X_0 = x > 0$ is an initial reserve and L_t^π is an increasing left-continuous process representing the cumulative dividends paid out by the company up till time t. Writing $\sigma^\pi = \inf\{t \geq 0 : \min\{U_t^\pi, U_{t+}^\pi\} < 0\}$ for the time at which ruin occurs, a dividend strategy is called admissible if, at any time before ruin, a lump sum dividend payment is smaller than the size of the available reserves: $L_{t^+}^\pi - L_t^\pi \leq U_t^\pi$ for $t \leq \sigma^\pi$.

The optimization objective function is given by the sum of the cumulative discounted dividends received until the moment of ruin and a negative payment or penalty that the beneficiaries of the dividends need to pay at ruin, which is a function of the shortfall at the moment $U_{\sigma^\pi}^\pi$ of ruin σ^π. More precisely,

$$v^\pi(x) = V^\pi(x) + H_w^\pi(x), \tag{3}$$

where

$$V^\pi(x) = \mathbb{E}_x \left[\int_0^{\sigma^\pi} \mathrm{e}^{-qt} \mathrm{d}L_t^\pi \right]$$

with rate of discounting $q > 0$, and H_w^π denotes the Gerber-Shiu penalty function

$$H_w^\pi(x) = \mathbb{E}_x \left[\mathrm{e}^{-q\sigma^\pi} w(U_{\sigma^\pi}) \right],$$

associated to a penalty $w : \mathbb{R} \to (-\infty, 0]$ (with $w(y) = 0$ for $y > 0$). In this article we will restrict ourselves to penalties that (on the negative half-axis) are equal to a constant multiple of the shortfall at the moment of ruin:

$$w(y) = cy, \qquad y < 0,\ c > 0.$$

The objective of the beneficiaries of the insurance company is to maximize $v^\pi(x)$ over all admissible strategies $\pi = \{L_t^\pi, t \geq 0\} \in \Pi$:

$$v_*(x) = \sup_{\pi \in \Pi} v^\pi(x), \tag{4}$$

where Π denotes the set of all admissible strategies.

Contents. In Section 2 the value function corresponding to a constant barrier strategy is expressed explicitly in terms of the scale function of the Crámer-Lundberg model. In Section 3 a verification theorem for the optimization problem (4) is given, which is used in Section 4 to analyze global optimality of the liquidation and constant barrier strategies. Finally, in Section 5 we present an analytically explicit solution to the case of exponential jumps.

§2. Barrier strategies and penalty functions

For a constant barrier strategy π_a at level a the value function $v_a = v^{\pi_a}$ is given by

$$v_a(x) = V^{\pi_a}(x) + H_w^{\pi_a}(x),$$

the sum of the present value V^{π_a} of the stream of dividend payments under the strategy π_a until the moment of ruin $\sigma^a = \sigma^{\pi_a} = \inf\{t \geq 0 : U_t^{\pi_a} < 0\}$ and the penalty at ruin

$$H_w^{\pi_a}(x) = \mathbb{E}_x[\mathrm{e}^{-q\sigma^a} w(U_{\sigma^a}^a)].$$

The functions V^{π_a} and H^{π_a} can be conveniently expressed in terms of the *q-scale function* $W^{(q)}$ of X that is defined as the unique continuous and increasing function $W^{(q)} : [0,\infty) \to [0,\infty)$ with the Laplace transform

$$\int_0^\infty \mathrm{e}^{-\theta y} W^{(q)}(y)\mathrm{d}y = (\psi(\theta) - q)^{-1}, \qquad \theta > \Phi(q), \tag{5}$$

where $\psi(\theta)$ is the Laplace-exponent of the Crámer-Lundberg process (1),

$$\psi(\theta) = p\theta - \lambda + \lambda \int_0^\infty \mathrm{e}^{-\theta y} F(\mathrm{d}y),$$

and $\Phi(q)$ is the positive root of the Crámer equation $\psi(\theta) = q$.
As shown in [3], V^{π_a} can be expressed in terms of $W^{(q)}$ as follows:

$$V^{\pi_a}(x) = \mathbb{E}_x\left[\int_0^{\sigma_a} \mathrm{e}^{-qt}\mathrm{d}L_t^a\right] = \frac{W^{(q)}(x)}{W^{(q)\prime}(a)}, \qquad x \in [0,a], \tag{6}$$

where $W^{(q)\prime}(a) = \lim_{x\downarrow a} W^{(q)\prime}(x)$ denotes the right limit at a ($W^{(q)\prime}$ has at most countably many discontinuities, at the locations of the atoms of F). Further, $L^a = L^{\pi_a}$ is the local time type strategy, given explicitly in terms of X by $L_0^a = 0$ and

$$L_t^a = \sup_{s\leq t} (X_s - a)^+, \qquad t > 0, \tag{7}$$

with $x^+ = \max\{x, 0\}$. Further, it was shown in Prop. 2 of [16] and [3] that under π_a, the Laplace transform of the ruin time and the expected discounted shortfall at ruin are given by

$$\mathbb{E}_x[\mathrm{e}^{-q\sigma^a}] = Z^{(q)}(x) - q\frac{W^{(q)}(a)}{W^{(q)\prime}(a)} W^{(q)}(x), \tag{8}$$

$$\mathbb{E}_x[\mathrm{e}^{-q\sigma^a} U_{\sigma^a}^a] = \overline{F}^{(q)}(x) - \frac{F^{(q)}(a)}{W^{(q)\prime}(a)}\, W^{(q)}(x), \tag{9}$$

where $U^a = U^{\pi_a}$ and $\overline{F}^{(q)}(x) = \int_0^x F^{(q)}(y)\mathrm{d}y$ with

$$F^{(q)}(x) = Z^{(q)}(x) - \psi'(0)W^{(q)}(x).$$

If the penalty is linear, $w(y) = cy$, $y < 0$, $-c < 0$, we thus arrive at the following result:

Proposition 1. *If $w(x) = cx$, $x < 0$, then:*
a) The penalty function $H_w^{\pi_a}(x)$, $x \in [0, a]$, is given by:

$$H_w^{\pi_a}(x) = \overline{F}_w(x) - W^{(q)}(x)\frac{F_w(a)}{W^{(q)\prime}(a)}, \tag{10}$$

where $\overline{F}_w(x) = \int_0^x F_w(y)\mathrm{d}y$ with

$$F_w(x) = cF^{(q)}(x). \tag{11}$$

b) The value function of a barrier strategy is given by:

$$v_a(x) = \overline{F}_w(x) + W^{(q)}(x)G_w(a), \qquad x \in [0, a], \tag{12}$$

where

$$G_w(a) = \frac{1 - F_w(a)}{W^{(q)\prime}(a)}. \tag{13}$$

§3. Verification theorem

To show the optimality of a particular dividend distribution strategy π across all admissible strategies Π for the dividend problem (4) we are led, by standard Markovian arguments, to consider the following variational inequalities:

$$\begin{aligned} \Gamma f(x) - qf(x) &\leq 0, \qquad x \geq 0, &(14)\\ f'(x) &\geq 1, \qquad x \geq 0, &(15)\\ f(x) &= cx, \quad x < 0, &(16) \end{aligned}$$

for functions $f : \mathbb{R} \to \mathbb{R}$ in the domain of the extended generator Γ of the process X which acts on C^1 functions f with compact support as

$$\Gamma f(x) = pf'(x) + \lambda \int_0^\infty [f(x - y) - f(x)]\, F(\mathrm{d}y), \tag{17}$$

where λ is the jump intensity and F the jump-size distribution. The following result, which is a special case of a general verification theorem proved in [4], implies that any sufficiently regular solution of the variational inequalities (14)–(16) dominates the value function v_* given in (4):

Theorem 2. *Suppose f is continuous and piecewise C^1 on $(0, \infty)$ and extend f to the negative half-line by setting $f(x) = cy$ for $y < 0$. If f satisfies (14) and (15), then $f \geq v_*$. If moreover there exists an admissible strategy $\pi \in \Pi$ such that $f = v_\pi$, then π is an optimal strategy and $v_* = v_\pi$.*

§4. Optimal strategies

If the penalty of the deficit at ruin is severe, it is clear that rather than continue the business it will be more profitable to liquidate the insurance company, by paying out all the initial reserves as a lump sum dividend payment, and force ruin immediately thereafter (by paying out dividends at a rate higher than the rate of premium income p).

Proposition 3 (Liquidation)**.** *Suppose $\mathcal{I}_c(x) \leq 0$ for all $x > 0$, where*

$$\mathcal{I}_c(x) := p - \lambda/\mu + \lambda(1-c)\int_{-\infty}^{0} \overline{F}(x-u)\mathrm{d}u - qx, \tag{18}$$

with $\overline{F}(x) = 1 - F(x)$ and $\mu^{-1} = \int_0^\infty xF(\mathrm{d}x)$ is the mean of F. Then it is optimal to liquidate the insurance company, for any level of the reserves.

Proof. Noting that the value function v_ℓ of the liquidation strategy is given by $v_\ell(x) = x$ for $x \geq 0$ and $v_\ell(x) = cx$ for $x < 0$ and using the form of Γ given in (17), it is a matter of straightforward calculation to show that, for $x > 0$, $(\Gamma v_l - qv_l)(x)$ is equal to the left-hand side of (18). The assertion then follows in view of Theorem 2. □

Let us next consider the complementary case that the ratio of premium income and expected payout per unit of time $p\mu/\lambda$ is sufficiently large compared to the slope c of the penalty function:

$$\mathcal{I}_c(0) = p - \lambda c/\mu > 0 \Leftrightarrow c < p\mu/\lambda. \tag{19}$$

Define the candidate optimal level a^* to be the (smallest) point where the function G_w attains its global maximum, where

$$G_w(a) = \frac{1 - F_w(a)}{W^{(q)\prime}(a)}, \tag{20}$$

that is,

$$a^* = \inf\{a \geq 0 : G_w(y) \leq G_w(a) \qquad \text{for all } y \geq 0\} \tag{21}$$

with $\inf \emptyset = \infty$. Under condition (19) the level a^* is finite. From Proposition 1 we see that the corresponding value function is given by

$$v_{a^*}(x) = v^{\pi_{a^*}}(x) = \begin{cases} \overline{F}_w(x) + W^{(q)}(x)G_w(a^*) & x \in [0, a^*] \\ x - a^* + v_{a^*}(a^*) & x > a^* \end{cases}. \tag{22}$$

Observing that $\mathcal{I}_w(x) \to -\infty$ as x tends to infinity, an argument similar to that used to prove the verification theorem can be employed to show that a^* is finite. Further, in view of the form (22) it follows that π_{a^*} is the optimal strategy among all constant barrier strategies, in the sense that

$$v_b(x) \leq v_{a^*}(x) \qquad \text{for all } x \in [0, a^*], b \geq 0.$$

We have the following sufficient conditions for global optimality of π_{a^*}:

Theorem 4. *Let condition* (19) *hold. The value function and optimal strategy of (4) are given by $v_* = v_{a^*}$ and $\pi_* = \pi_{a^*}$ if (i)*

$$\Gamma v_{a^*} - qv_{a^*}(x) \leq 0 \qquad \textit{for all } x > a^* \tag{23}$$

or (ii) G_w is monotone decreasing on (a^*, ∞), i.e.

$$G_w(a) \geq G_w(b), \qquad \textit{for } a^* < a < b. \tag{24}$$

Let us sketch the proof of this result (see also [4] for a more general version of this result). The idea of the proof is to show first that v_{a^*} satisfies the HJB equation (14) – (16) and to apply next the verification theorem, Theorem 2. For statement (i) we note to that end that

(a) Since
$$\left\{ \mathrm{e}^{-q(t\wedge T_0^-)} \left(W^{(q)}(X_{t\wedge T_0^-}) G(a^*) + \overline{F}^{(q)}(X_{t\wedge T_0^-}) \right), t \geq 0 \right\}$$
is a martingale it holds that $\Gamma v_{a^*}(x) - q v_{a^*}(x) = 0$ for $x \in (0, a^*)$ (see [4, Lemma 10]).

(b) For $x \in (0, a^*)$ it holds that
$$v'_{a^*}(x) \geq 1,$$
which is a consequence the form of v_a and the facts that $G_c(x) \leq G_c(a^*)$ for all $0 < x \leq a^*$ and $v'_x(x) = 1$

(ii) By arguments similar to those employed in the proof of Theorem 2 of Loeffen [14] it can be shown that

$$\Gamma v_{a^*} - q v_{a^*}(x) \leq 0 \text{ for } x > a^* \text{ if } v'_a(x) \geq v'_b(x) \text{ for all } x \text{ and } a, b \text{ with } a^* < a < b,$$

which is equivalent to (24) in view of the form (12) of v_{a^*}. The assertion follows then in view of (i).

In the next section we will use the condition (24) to explicitly and completely analyse the case of exponential jumps.

§5. Cramér-Lundberg model with exponential jumps

Suppose X is given by the Cramér-Lundberg model (1) with exponential jump sizes with mean $1/\mu$. Let $p > \lambda/\mu$. Then $\kappa(\theta) = p\theta - \lambda\theta/(\mu + \theta)$ and the scale functions $W^{(q)}$ and $Z^{(q)}$ are given by

$$\begin{aligned} W^{(q)}(x) &= p^{-1} \left(A_+ \mathrm{e}^{q^+(q)x} - A_- \mathrm{e}^{q^-(q)x} \right), \\ Z^{(q)}(x) &= p^{-1} q \left(q^+(q)^{-1} A_+ \mathrm{e}^{q^+(q)x} - q^-(q)^{-1} A_- \mathrm{e}^{q^-(q)x} \right) \\ &= \mu^{-1} \left(q^+(q) A_- \mathrm{e}^{q^-(q)x} - q^-(q) A_+ \mathrm{e}^{q^+(q)x} \right), \end{aligned}$$

where $A_\pm = \frac{\mu + q^\pm(q)}{q^+(q) - q^-(q)}$ with $q^+(q) = \Phi(q)$ and $q^-(q)$ the smallest root of $\kappa(\theta) = q$:

$$q^\pm(q) = \frac{q + \lambda - \mu\, p \pm \sqrt{(q + \lambda - \mu\, p)^2 + 4pq\mu}}{2p}.$$

We recall from [3] that in the absence of the penalty ($w(x) = 0$), the function $G(x)^{-1} = W^{(q)\prime}(x)$ is unimodular with global minimum at

$$a^* = \frac{1}{q^+(q) - q^-(q)} \begin{cases} \log \frac{q^-(q)^2(\mu + q^-(q))}{q^+(q)^2(\mu + q^+(q))} & \text{if } W^{(q)\prime\prime}(0) < 0 \Leftrightarrow (q+\lambda)^2 - p\lambda\mu < 0 \\ 0 & \text{if } W^{(q)\prime\prime}(0) \geq 0 \Leftrightarrow (q+\lambda)^2 - p\lambda\mu \geq 0 \end{cases}$$

(since $W^{(q)\prime\prime}(0) \sim q^+(q)^2(\mu + q^+(q)) - q^-(q)^2(\mu + q^-(q))/(q^+(q)) - q^-(q)) = (q + \lambda)^2 - p\lambda\mu$ and therefore the optimal strategy is always the barrier strategy at level a^*.
We take now the linear penalty function $w(x) = cx$. From Proposition 3, liquidation is optimal for any level of reserves, if $\mathcal{I}_w(0) \leq 0 \Leftrightarrow c \geq \rho^{-1} = p\mu/\lambda$.
We will show now that in the opposite case $0 < c < \rho^{-1}$, the optimal strategy is a barrier strategy at some level a^*.
Let us investigate the sign of the function $D(x) = -G'(x)W^{(q)\prime}(x)$,

$$\begin{aligned} &D(x) := W^{(q)\prime\prime}(x)(1 - cF(x)) + cF^{(q)\prime}(x)W^{(q)\prime}(x) = W^{(q)\prime\prime}(x) \\ &+c\left[W^{(q)\prime\prime}(x)(-Z^{(q)}(x) + \kappa'(0)W^{(q)}(x))) + W^{(q)\prime}(x)(qW^{(q)}(x) - \kappa'(0)W^{(q)\prime}(x))\right], \end{aligned}$$

which determines the optimal barrier policy.
If a positive root $a^* > 0$ of $D(x)$ exists, it must satisfy:

$$\begin{aligned} &W^{(q)\prime\prime}(a^*) + cW^{(q)\prime\prime}(a^*)(\kappa'(0)W^{(q)}(a^*) - Z^{(q)}(a^*)) \\ &+ \quad cW^{(q)\prime}(a^*)(qW^{(q)}(a^*) - \kappa'(0)W^{(q)\prime}(a^*)) = 0. \end{aligned} \tag{25}$$

Noting that

$$\begin{aligned} W^{(q)\prime\prime}(x)W^{(q)}(x) - (W^{(q)\prime}(x))^2 &= -A_+A_-(q^+ - q^-)^2 \mathrm{e}^{(q^+ + q^-)x} \\ W^{(q)\prime}(x)W^{(q)}(x) - W^{(q)\prime\prime}(x)q^{-1}Z^{(q)}(x) &= P\mathrm{e}^{(q^+ + q^-)x}, \end{aligned}$$

where

$$P = A_+A_-(q^+ - q^-)^2(q^+ + q^-)/(q^+ q^-) = (\mu + q^+)(\mu + q^-)(\kappa'(0)/q - 1/\mu)$$

and where we used $(q^+ + q^-)/(q^+ q^-) = \kappa'(0)/q - 1/\mu$, we see that (25) is equivalent to

$$cA\mathrm{e}^{(q^+(q)+q^-(q))a^*} - B\mathrm{e}^{q^+(q)a^*} + C\mathrm{e}^{q^-(q)a^*} = 0,$$

where we changed the sign and where

$$\begin{aligned} A &= \frac{q}{\mu}(\mu + q^+(q))(\mu + q^-(q)) > 0, \\ B &= q^+(q)^2 A_+ > 0, \\ C &= q^-(q)^2 A_- > 0. \end{aligned}$$

We note now that $a^* = 0$ iff $(q+\lambda)^2 - \lambda\mu p \geq c\lambda q \Leftrightarrow G'(0) \leq 0$. To verify this note that

$Z^{(q)}(0) = 1, W^{(q)}(0) = p^{-1}, W^{(q)'}(0) = p^{-2}(q+\lambda)$ and $W^{(q)''}(0) = p^{-3}[(q+\lambda)^2 - \lambda\mu p]$.

Thus, $G'(0) \leq 0$ is equivalent to $c\rho(q+\lambda)(p\mu-(q+\lambda))+(c\rho-1)((q+\lambda)^2-\lambda p\mu) \leq 0$. In conclusion, from Theorem 4 it follows that the barrier strategy at level a^* is globally optimal if

$$f(x) := cA\mathrm{e}^{q^+(q)x} - B\mathrm{e}^{(q^+(q)-q^-(q))x} + C < 0 \qquad \text{for } x > a^*. \tag{26}$$

To see that this is the case note that $f(x) \downarrow +C$ and $\downarrow -\infty$ as $x \to \pm\infty$ and that f attains its maximum at unique $x^* \in \mathbb{R}$, with $f'(x) > 0$ and $f'(x) < 0$ for $x < x^*$ and $x > x^*$. Since $f(x^*) > C$, we deduce that $x^* < a^*$, and (26) holds true.

Acknowledgements

This work is supported by EPSRC grant EP/D039053 and the Ministry of Science and Higher Education of Poland under the grant N N201 394137 (2009-2011).

References

[1] ALBRECHER, H. AND THONHAUSER, S. (2007) Dividend maximization under consideration of the time value of ruin, *Insurance: mathematics and economics* **41**, pp. 163–184.

[2] ASMUSSEN, S., HØJGAARD, B. AND TAKSAR, M. (2000) Optimal risk control and dividend distribution policies. Example of excess-of loss reinsurance for an insurance corporation, *Finance Stoch.* **4**, pp. 299–324.

[3] AVRAM, F., PALMOWSKI, Z. AND PISTORIUS, M.R. (2007) On the optimal dividend problem for a spectrally negative Lévy process, *Ann. Appl. Probab.* **17**, pp. 156–180.

[4] AVRAM, F., PALMOWSKI, Z. AND PISTORIUS, M. (2008) On optimal dividend distribution for a Lévy risk-process in the presence of a Gerber-Shiu penalty function. *Manuscript.*

[5] DICKSON, D.C.M. AND WATERS, H.R. (2004) Some optimal dividends problems, ASTIN Bulletin, 34: 49–74

[6] DE FINETTI, B. (1957) Su un'impostazione alternativa dell teoria colletiva del rischio, *Trans. XV Intern. Congress Act.* **2**, pp. 433–443.

[7] GERBER, H.U. (1969) Entscheidungskriterien für den Zusammengesetzten Poisson Prozess, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **69**, pp. 185–228.

[8] GERBER, H.U. (1972) Games of Economic Survival with Discrete- and Continuous-Income Processes, *Operations Research* **20**, pp. 37–45.

[9] GERBER, H.U. (1979) *An introduction to mathematical risk theory,* Hübner Foundation for Insurance Education series 8, Philadelphia.

[10] GERBER, H.U. AND SHIU, E.S.W. (2004) Optimal dividends: analysis with Brownian motion, *North American Actuarial Journal* **8**, pp. 1–20.

[11] GERBER, H.U., LIN, X.S. AND YANG, H. (2006) A note on the Dividends-Penalty Identity and the Optimal Dividend Barrier, *Astin Bull.* **36**, pp. 489Ű-503.

[12] HALLIN, M. (1979) Band strategies: The random walk of reserves, *Blätter der DGVFM* **14**, pp. 321–236.

[13] JEANBLANC, M. AND SHIRYAEV, A.N. (1995) Optimization of the flow of dividends, *Russian Math. Surveys* **50**, pp. 257–277.

[14] LOEFFEN, R. (2008) On optimality of the barrier strategy in de Finetti's dividend problem for spectrally negative Lévy processes, *Annals of Applied Probability* **18(5)**, pp. 1669–1680.

[15] LOEFFEN, R. AND RENAUD, J-F. (2009) De Finetti's optimal dividend problem with an affine penalty function at ruin. *I. M. E.* to appear.

[16] PISTORIUS, M.R. (2004) On exit and ergodicity of the completely asymmetric Lévy process reflected at its infimum, *J. Th. Probab.* **17**, pp. 183–220.

[17] YUEN, K.C., WANG, G. AND LI, W.K. (2006) The Gerber-Shiu expected discounted penalty function for risk processes with interest and a constant dividend barrier, *Insurance: Mathematics and Economics* **40(1)**, pp. 104–112.

[18] ZAJIC, T. (2000) Optimal Dividend Payout under Compound Poisson Income, *J. Opt. Th. Appl.* **104**, pp. 195–213.

[19] ZHOU, X. (2005) On a classical risk model with a constant dividend barrier, *North American Actuarial Journal* **9**, pp. 1–14.

Florin Avram
University of Pau (France)
florin.avram@univ-pau.fr

Martijn Pistorius
Department of Mathematics
King's College London (United Kingdom)
Martijn.Pistorius@kcl.ac.uk

Zbigniew Palmowski
Department of Mathematics
University of Wroclaw (Poland)
zpalma@math.uni.wroc.pl

Monografías Matemáticas García de Galdeano **36**, 79–88 (2010)

A BIVARIATE STOCHASTIC GAMMA DIFFUSION MODEL: STATISTICAL INFERENCE

R. Gutiérrez, R. Gutiérrez-Sánchez, A. Nafidi

Abstract. In the present study, we propose a bivariate stochastic Gamma diffusion model as the solution to Ito's stochastic differential equations (SDE) that are similar as regards the drift and diffusion coefficients to those considered in the univariate Gamma diffusion model (see [11]). Firstly, we determine the main probabilistic characteristics of this model, such as the solution to the SDE, the bivariate transition density, the bidimensional moment functions, the conditioned trend functions and in particular, the correlation function between each of the components of the model. Then, based on some results of matrix differential calculus (see [13]), the statistical inference in the model is drawn, estimating the bidimensional drift and the diffusion matrix by the maximum likelihood method using discrete sampling. Finally, we obtain the properties of the resulting likelihood estimators.

Keywords: Bivariate stochastic Gamma diffusion process, Likelihood estimation using discrete sampling, Matrix differential calculus, Normal and Wishart random matrices.

AMS classification: 60J60, 62M05.

§1. Introduction

Stochastic diffusion processes are of great interest to investigators in many fields, such as biology, physics, demography, economics and environmental sciences. One of the questions that has aroused greatest interest about these stochastic models (especially in the one dimensional case) and one that has been the object of numerous studies in recent years, is that of statistical estimation and inference. Various methods addressing the question of statistical inference have been developed recently, and several papers published on the topic, including those by Biby and Sorensen [4] and Ait-Sahalia [1], without overlooking the wide-ranging review of results presented by Prakasa Rao [14], who provides a lengthy list of references on the subject.

As regards the statistical inference in multivariate diffusion processes, some special cases have been studied, such as the multivariate lognormal and Gompertz diffusion processes. An extensive study of the probabilistic aspects and of the corresponding statistical inference (estimation and test of hypotheses) can be seen, for example, for the first process in [6], [7] and [8] and for the second process in [9] and [5].

The main aim of this study is to propose and examine a "bivariate Gamma diffusion model". To do so, we follow the methodology previously employed to extend one-dimensional diffusion processes to the multivariate case. For example, in [6] and [7] this methodology is applied to the case of a multivariate lognormal diffusion model, and in [9] for the case of the bivariate Gompertz diffusion model. Specifically, in the present study, we extend the univariate stochastic Gamma diffusion process studied in [10] and [11] to a bivariate Gamma process. The latter process is constituted of two components, which are Gamma univariate processes and which are interrelated in the sense that they vary in a correlated way in their stochastic evolution in time. Having established this model, we then study its probabilistic characteristics and its associated basic statistical inference.

This paper is organized as follows. In the second section, we determine the main probabilistic characteristics of the model, such as the solution to the SDE, the bivariate transition density, the bidimensional moment functions, the conditioned trend functions and in particular, the correlation function between each of the components of the model. In the third section, the statistical inference in the model is achieved, and the bidimensional drift and the diffusion matrix are estimated by the maximum likelihood method based on discrete sampling. In the last section, based on some results of the matrix normal distribution, we obtain the properties of the resulting likelihood estimators.

§2. Bivariate SGDP and its characteristics

2.1. The model and its analytical expression

Let $\{x(t) = (x_1(t), x_2(t))'; t \in [t_0, T]; t_0 > 0\}$ be a bivariate stochastic process that satisfies the following Ito's SDE:

$$dx(t) = a(t, x(t))dt + b(t, x(t))dw(t) \quad ; \quad x(t_0) = x_{t_0} \tag{1}$$

with the vector $a(t, x)$ and the matrix $b(t, x)$ being given as follows

$$a(t, x) = D(x)\left(\frac{a}{t} - \beta\right) \quad ; \quad b(t, x) = D(x)B^{1/2}$$

where $\{w(t); t \in [t_0, T]\}$ is a 2-dimensional standard Wiener process, x_{t_0} is a fixed vector belonging to $(0, \infty)^2$, $x = (x_1, x_2)' \in (0, \infty)^2$, $a = (a_1, a_2)'$ and $\beta = (\beta_1, \beta_2)'$, in which $D(x)$ is a diagonal matrix where the elements of the principal diagonal are x_1, x_2, and where $B = (b_{ij})_{i,j}$ is a 2×2 symmetric non negative definite matrix. The parameters $a_1, a_2, \beta_1, \beta_2$ and $b_{i,j}$ for $1 \leq i, j \leq 2$ are real and will be the object of subsequent statistical estimation.

The vector $a(t, x)$ and the matrix $b(t, x)$ specified in Eq.(2) satisfy the Lipschitz restriction on growth conditions for the existence and unicity of the solution to the SDEs in theorem (6.2.2 page 105) of Arnold [3]. Thus, let:

$$K_1 = \max_{1 \leq i \leq 2} \left[\max_{t \in [t_0, T]} \left(\frac{a_i}{t} - \beta_i \right)^2 \right] \quad \text{and} \quad K_2 = \max_{1 \leq i \leq 2} (b_{ii}).$$

Then, on the one hand, $\exists K = K_1^{1/2} + K_2^{1/2}$, $\forall t \in [t_0, T]$, such that $\forall x, y \in (0, \infty)^2$ we have

$$\|a(t,x) - a(y,t)\|_e + \|b(t,x) - b(t,y)\|_{tr} \leq K\|x - y\|_e$$
$$\|a(t,x)\|_e^2 + \|b(t,x)\|_{tr}^2 \leq K(1 + \|x\|_e^2)$$

where $\| \quad \|_e$ denotes the euclidean norm in $\mathbb{R}^2$ and $\| \quad \|_{tr}$ denotes the trace norm in $\mathcal{M}_{2\times 2}$ ($\|A\|_{tr} = [tr(AA')]^{1/2}$).

Then, under these conditions equation Eq.(1) has on $[t_0, T]$ a unique $\mathbb{R}^2$-valued solution $\{x(t); t \in [t_0, T]\}$, continuous with probability 1, that satisfies the initial condition $x(t_0) = x_{t_0}$.

On the other hand, as the vector $a(t,x)$ is a continuous function with respect to t, then by theorem (9.3.1 page 152) of Arnold [3], the solution $\{x(t); t \in [t_0, T]\}$ is a 2-dimensional diffusion process on $[t_0, T]$ with drift vector $a(t,x)$ and with a diffusion matrix given by

$$B(x) = (D(x)B^{1/2})(D(x)B^{1/2})' = D(x)BD(x) = (b_{ij}x_ix_j)_{1\leq i,j\leq 2}\,.$$

The analytical expression of process $\{x(t), t \in [t_0, T]\}$ can be obtained by applying Ito's formula (see, for example [3]) to a transform of the type $y(t) = \log(x(t)) = (\log(x_1(t)), \log(x_2(t)))'$, and then we obtain

$$dy(t) = \left[\frac{a}{t} - \left(\beta + \frac{b}{2}\right)\right] dt + B^{1/2} dw(t) \quad , \quad y(t_0) = \log(x_{t_0})$$

where $b = (b_{11}, b_{22})'$, and then by integration we have

$$y(t) = y(t_0) + \log\left(\frac{t}{t_0}\right) a - \left(\beta + \frac{b}{2}\right)(t - t_0) + B^{1/2}(w_t - w_{t_0})$$

from which we can deduce that the solution to the original SDE Eq. (1) has the following form

$$x(t) \quad = \quad \exp\left(\log(x_{t_0}) + \log\left(\frac{t}{t_0}\right) a - \left(\beta + \frac{b}{2}\right)(t - t_0) + B^{1/2}(w_t - w_{t_0})\right).$$

2.2. The ptdf and moments of the model

Taking into account that the random vector $(w(t) - w(s))$ has a bivariate normal distribution $\mathcal{N}_2\,(0, (t-s)I_2)$ (where I_2 denotes the 2×2 identity matrix), it can be deduced that $x(t) \mid x(s) = x_s$ has a bivariate lognormal distribution $\Lambda_2\,(\mu(s,t,x_s), (t-s)B)$ where $\mu(s,t,x_s)$ is the following 2-dimensional vector

$$\mu(s,t,x) \quad = \quad \log(x) + a\log\left(\frac{t}{s}\right) - \left(\beta + \frac{b}{2}\right)(t-s) \tag{2}$$

and therefore the transition density function of the process $f(y,t \mid x,s)$ (for $y = (y_1,y_2)'$ and $x = (x_1,x_2)'$) has the form

$$f(y,t \mid x,s) = [2\pi]^{-1}(t-s)^{-1} \mid B \mid^{-\frac{1}{2}} (y_1y_2)^{-1} \exp\left\{-\frac{Q}{2}\right\} \tag{3}$$

where $\mid B \mid$ is the determinant of the matrix B, and Q is a quadratic form that is given by

$$Q \quad = \quad (\log(y) - \mu(s,t,x))' \left[(t-s)B\right]^{-1} (\log(y) - \mu(s,t,x))$$

where $\mu(s,t,x)$ is as given in Eq. (2).

The marginal conditional and non-conditional moments of order r ($r \in \mathbb{N}^*$) can be obtained from the function generating the random vector $Z(t) = \log\left[x(t) \mid x(s) = x_s\right]$, which follows the law $\mathcal{N}_2\left(\mu(s,t,x_s);(t-s)B\right)$, and is expressed as follows, for $\lambda \in \mathbb{R}^2$

$$\mathbb{E}(e^{\lambda' Z(t)}) \quad = \quad \exp\left\{\lambda'\mu(s,t,x_s) + \frac{t-s}{2}\lambda' B\lambda\right\}.$$

For particular values of the vector $\lambda = (0,r)'$ or $\lambda = (r,0)'$ ($r \in \mathbb{N}^*$), we obtain, for example, the marginal conditional trend functions of order r of the process and which have the following form, for $i = 1,2$

$$\mathbb{E}\left(x_i^r(t) \mid x_i(s) = x_{s,i}\right) \quad = \quad \exp\left(r\mu_i(s,t,x_s) + \frac{r^2(t-s)}{2}b_{ii}\right) \tag{4}$$

and for $\lambda = (r_1,r_2)'$ ($r_1,r_2 \in \mathbb{N}^*$), we obtain the joint conditional trend of the process

$$\begin{aligned}\mathbb{E}\left(x_1^{r_1}(t)x_2^{r_2}(t) \mid x(s) = x_s\right) \quad &= \quad \exp\left(r_1\mu_1(s,t,x_s) + r_2\mu_2(s,t,x_s)\right. \\ &\quad \left. + \frac{(t-s)}{2}(r_1^2b_{11} + r_2^2b_{22} + 2r_1r_2b_{12})\right).\end{aligned} \tag{5}$$

Using Eq. (4) in the particular case $r = 1$, we obtain the marginal conditional trend function of the process

$$\mathbb{E}\left(x_i(t) \mid x_i(s) = x_{s,i}\right) \quad = \quad \exp\left(\mu_i(s,t,x_s) + \frac{1}{2}(t-s)b_{ii}\right). \tag{6}$$

By assuming the initial condition $\mathrm{P}(x(t_0) = x_{t_0}) = 1$, and using Eq. (6) then the non conditional marginal trend functions are

$$\begin{aligned}\mathbb{E}\left(x_i(t)\right) &= \exp\left(\mu_i(t_0,t,x_{t_0}) + \frac{1}{2}(t-t_0)b_{ii}\right)\\ &= \frac{x_{t_0,i}e^{\beta_i t_0}}{t_0^{a_i}} t^{a_i} e^{-\beta_i t}.\end{aligned}$$

From Eq. (4) and Eq. (6), we can deduce that the marginal variance function of the process, for $i = 1, 2$ is:

$$\mathrm{Var}\left(x_i(t)\right) = \exp\left(2\mu_i(t_0,t;x_{t_0}) + (t-s)b_{ii}\right)\left(e^{(t-s)b_{ii}} - 1\right)$$

and the covariance function at the same instant is

$$\begin{aligned}\mathrm{Cov}\left(x_1(t),x_2(t)\right) &= \exp\left(\mu_1(t_0,t;x_{t_0}) + \mu_2(t_0,t;x_{t_0}) + \frac{1}{2}(t-s)(b_{11}+b_{22})\right)\\ &\quad \left(e^{(t-s)b_{12}} - 1\right).\end{aligned}$$

The correlation function of the process at the same instant is given by

$$\varrho(x_1(t),x_2(t)) = \frac{\left(e^{(t-t_0)b_{12}} - 1\right)}{\left(e^{(t-t_0)b_{11}} - 1\right)^{1/2}\left(e^{(t-t_0)b_{22}} - 1\right)^{1/2}}.$$

§3. Statistical inference on the model

3.1. Parameter likelihood estimation

Let us now obtain the maximum likelihood estimators of the parameters corresponding to the model β, a and B, using discrete sampling. To construct the likelihood function associated with the process, the following discrete sampling is used: $\{x(t_1) = x_{t_1}; x(t_2) = x_{t_2}; \ldots, x(t_n) = x_{t_n}\}$ at the instants $t_1, t_2; \ldots; t_n$, in which each $x(t_\alpha)$ represents the bidimensional vector $x(t_\alpha) = (x_1(t_\alpha), x_2(t_\alpha))'$, which for the sake of simplicity we shall denote as $x_{t_\alpha} = x_\alpha$. We also considered the initial condition $\mathrm{P}[x(t_1) = x_1] = 1$; by applying the Markov property and making use of Eq. (3), the likelihood function associated with the sample considered, of size $(n-1)$ is given by

$$\begin{aligned}&\mathbb{L}(x_1,\ldots,x_n) = \prod_{\alpha=2}^{n} f(x_\alpha, t_\alpha \mid x_{\alpha-1}, t_{\alpha-1})\\ &= (2\pi)^{-\frac{k(n-1)}{2}} \mid B \mid^{-\frac{(n-1)}{2}} \prod_{\alpha=2}^{n}(t_\alpha - t_{\alpha-1})\left(\prod_{i=1}^{k} x_{\alpha,i}^{-1}\right)\\ &\exp\left\{-\frac{1}{2}\left[\log(x_\alpha/x_{\alpha-1}) - a\log(t_\alpha/t_{\alpha-1}) + \left(\beta + \frac{b}{2}\right)(t_\alpha - t_{\alpha-1})\right]'\right.\\ &\left.(t_\alpha - t_{\alpha-1})^{-1}B^{-1}\left[\log(x_\alpha/x_{\alpha-1}) - a\log(t_\alpha/t_{\alpha-1}) + \left(\beta + \frac{b}{2}\right)(t_\alpha - t_{\alpha-1})\right]\right\}.\end{aligned}$$

By carrying out the following change of variable: $v_1 = x_1$ and $v_\alpha = (t_\alpha - t_{\alpha-1})^{-1/2} (\log(x_\alpha) - \log(x_{\alpha-1}))$ for $\alpha = 2, \ldots, n$, then, in terms of v_α, the likelihood function is given by

$$\begin{aligned} \mathbb{L}_{\mathbf{v}_1,\ldots,\mathbf{v}_n}(\Gamma; B) &= (2\pi)^{-(n-1)k/2} \mid B \mid^{-\frac{(n-1)}{2}} \\ &\quad \exp\left\{ -\frac{1}{2} \sum_{\alpha=2}^{n} (\mathbf{v}_\alpha - \Gamma u_\alpha)' B^{-1} (\mathbf{v}_\alpha - \Gamma u_\alpha) \right\} \end{aligned} \tag{7}$$

where, $u_\alpha = (t_\alpha - t_{\alpha-1})^{-1/2} (\log(t_\alpha/t_{\alpha-1}), t_\alpha - t_{\alpha-1})'$, for $\alpha = 2, \ldots, n$, and $\Gamma = \left(a, -\left(\beta + \frac{b}{2}\right)\right)$ is (2×2)- matrix.

Let $\mathbf{V} = (v_2, \ldots, v_n)$ and $\mathbf{U} = (u_2, \ldots, u_n)$. The likelihood function can then be written as follows:

$$\mathbb{L}_{\mathbf{V}} = (2\pi)^{-(n-1)} \mid B \mid^{-\frac{(n-1)}{2}} \exp\left\{ -\frac{1}{2} tr\left[B^{-1} (\mathbf{V} - \Gamma\mathbf{U}) (\mathbf{V} - \Gamma\mathbf{U})' \right] \right\}.$$

By taking the logarithm, we obtain

$$\log(\mathbb{L}_{\mathbf{V}}) = -(n-1)\frac{k}{2}\log(2\pi) - \frac{n-1}{2}\log \mid B \mid -\frac{1}{2} tr\left[B^{-1} (\mathbf{V} - \Gamma\mathbf{U}) (\mathbf{V} - \Gamma\mathbf{U})' \right].$$

Then, calculating the differential of this function, and making use of the following results of matrix differential calculus (see, for example [13]): $d\left[\text{tr}(B)\right] = \text{tr}\,(dB)$, $d\left[\log \mid B \mid\right] = \text{tr}\left(B^{-1}dB\right)$ and $d\left[B^{-1}\right] = B^{-1}(dB)B^{-1}$, we have

$$\begin{aligned} d\log(\mathbb{L}) &= -\frac{n-1}{2}\text{tr}\left(B^{-1}dB\right) - \frac{1}{2} tr\left[-B^{-1}(dB)B^{-1} (\mathbf{V} - \Gamma\mathbf{U}) (\mathbf{V} - \Gamma\mathbf{U})' \right] \\ &\quad -\frac{1}{2} tr\left[B^{-1} (-d\Gamma)\, \mathbf{U} (\mathbf{V} - \Gamma\mathbf{U})' + B^{-1} (\mathbf{V} - \Gamma\mathbf{U})\, \mathbf{U}' (-d\Gamma)' \right]. \end{aligned}$$

By applying trace properties, the above differential can be written as follows:

$$\begin{aligned} d\log(\mathbb{L}) &= \frac{1}{2}\text{tr}\left\{ \left[B^{-1} (\mathbf{V} - \Gamma\mathbf{U}) (\mathbf{V} - \Gamma\mathbf{U})' - (n-1)I_2 \right] B^{-1}dB \right\} \\ &\quad + tr\left\{ \mathbf{U} (\mathbf{V} - \Gamma\mathbf{U})' B^{-1} d\Gamma \right\}. \end{aligned}$$

From the relations $tr(AB) = Vec'(A')Vec(B)$ and $dVec(A) = Vec(dA)$, where Vec denotes the matrix vectorization (given an $n \times m$ matrix X, the $Vec(X)$ is the vector of dimension $nm \times 1$ that stacks the columns of X), we obtain

$$\begin{aligned} d\log(\mathbb{L}) &= \frac{1}{2} Vec'\left\{ \left[B^{-1} (\mathbf{V} - \Gamma\mathbf{U}) (\mathbf{V} - \Gamma\mathbf{U})' - (n-1)I_2 \right] B^{-1} \right\} d\text{Vec}(B) \\ &\quad + Vec'\left\{ B^{-1} (\mathbf{V} - \Gamma\mathbf{U})\, \mathbf{U}' \right\} d\text{Vec}(\Gamma). \end{aligned}$$

Then, making this differential equal to zero, with respect to the estimators of B and γ, we obtain

$$B^{-1}\left(\mathbf{V}-\Gamma\mathbf{U}\right)\mathbf{U}'=0 \tag{8}$$

$$\left[B^{-1}\left(\mathbf{V}-\Gamma\mathbf{U}\right)\left(\mathbf{V}-\Gamma\mathbf{U}\right)'-(n-1)I_2\right]B^{-1}=0. \tag{9}$$

From Eq. (8) and Eq. (9), we obtain the maximum likelihood estimators of the matrices Γ and B, which are given by

$$\widehat{\Gamma}=\mathbf{V}\mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1} \tag{10}$$

$$(n-1)\widehat{B}=\mathbf{V}\mathbf{H}_{\mathbf{U}}\mathbf{V}' \tag{11}$$

where $\mathbf{H}_{\mathbf{U}}=I_{n-1}-\mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1}\mathbf{U}$ is an idempotent symmetric matrix.

3.2. Likelihood estimator distribution

In order to study the estimator distributions obtained by the expressions Eq. (10) and Eq. (11), it is necessary to examine some results of the matrix normal distribution (see, for example [13]), which are presented as follows:

Definition 1. : Let $X_{m\times n}$ be a random matrix and let $M_{m\times n}$, $C_{m\times m}$ and $D_{n\times n}$ be constant matrices (C and D are non negative definite matrices). We then say that the random matrix X has a normal distribution and it is denoted by $\mathcal{N}_{m\times n}\left(M;C\otimes D\right)$ ($\otimes$ denotes the Kronecker product of matrices: $C\otimes D=[c_{ij}D]$) if the density function of X is

$$f(x)=(2\pi)^{-\frac{mn}{2}}\mid C\mid^{-\frac{m}{2}}\mid D\mid^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}tr\left[C^{-1}\left(M-x\right)D^{-1}\left(M-x\right)'\right]\right\}.$$

In the sense of the matrix vectorization, we have the following equivalence:

$$X\sim\mathcal{N}_{m\times n}\left(M;C\otimes D\right)\text{ if only if }vec(X)\sim\mathcal{N}_{mn}\left(vec(M');C\otimes D\right).$$

Corollary 1. *: Let $X\sim\mathcal{N}_{m\times n}\left(M;C\otimes D\right)$ and let $N_{p\times m}$ be a constant matrix. We then have*

$$X'\sim\mathcal{N}_{n\times m}\left(M';D\otimes C\right).$$

$$NX\sim\mathcal{N}_{n\times p}\left(NM;NCN'\otimes D\right).$$

3.2.1. Distribution of $\widehat{\Gamma}$

The expression in Eq.(7) can be rewritten as follows

$$\mathbb{L}_{\mathbf{V}}=(2\pi)^{-(n-1)}\mid B\mid^{-\frac{n-1}{2}}\mid I_{n-1}\mid^{-1}\exp\left\{-\frac{1}{2}tr\left[B^{-1}\left(\mathbf{V}-\Gamma\mathbf{U}\right)I_{n-1}^{-1}\left(\mathbf{V}-\Gamma\mathbf{U}\right)'\right]\right\}.$$

From which we deduce the matrix

$$\mathbf{V}\sim\mathcal{N}_{2\times(n-1)}\left(\Gamma\mathbf{U};B\otimes\mathbf{I}_{n-1}\right).$$

Then, by using Corollary 1, we have

$$\mathbf{V}\mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1} \sim \mathcal{N}_{2\times 2}\left(\Gamma\mathbf{U}\mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1}; B\otimes\left(\mathbf{U}\mathbf{U}'\right)^{-1}\mathbf{U}\mathrm{I}_{n-1}\mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1}\right).$$

Thus, we obtain that

$$\widehat{\Gamma} \sim \mathcal{N}_{2\times 2}\left(\Gamma, B\otimes\left(\mathbf{U}\mathbf{U}'\right)^{-1}\right).$$

3.2.2. Distribution of $\widehat{B}$

To obtain the distribution of the matrix $\widehat{B}$, we make use of the following result (see for example [15], corollary 3.2):

Corollary 2. *If $Y \sim \mathcal{N}_{n\times p}\left[\mu, A\otimes\Sigma\right]$, then $Y'WY$ has a non central Wishart distribution with m degrees of freedom, covariance Σ and noncentral matrix λ noted by $\mathcal{W}_p(m,\Sigma,\lambda)$, if and only if:*

$$AWAWA = AWA, \qquad tr(AW) = m$$
$$\lambda = \mu'W\mu = \mu'WAW\mu = \mu'WAWAW\mu.$$

Using the latter result in the particular case: $Y = V'$, $A = \mathrm{I}_{n-1}$, $\Sigma = B$, $W = \mathrm{H}_{\mathbf{U}}$ and $\mu = U'\Gamma'$, and so we have: $\mathrm{tr}(AW) = m = n-3$ and $\lambda = 0$, and $V\mathrm{H}_{\mathbf{U}}V' \sim \mathcal{W}_2(n-3, B)$ and therefore by symmetric properties of the Wishart distribution, we deduce that

$$(n-1)\hat{B} \sim \mathcal{W}_2(n-3, B).$$

3.2.3. Independence of likelihood estimators

To show that $\widehat{\Gamma}$ and $\hat{B}$ are independently distributed, we make use of the following independence result between linear and quadratic forms (see for example [12] corollary 6):

Corollary 3. *Let $Y \sim \mathcal{N}_{n\times p}\left[\mu, A\otimes\Sigma\right]$, Then, the necessary and sufficient conditions for the independence of $YWY' + \frac{1}{2}\left(LY' + YL' + C\right)$ and YM' are $AWM' = 0$ and $LWM' = 0$.*

By applying this result to the particular case $Y = V$, $W = \mathrm{H}_{\mathbf{U}}$, $A = B$, $\Sigma = I_{n-1}$, $L = 0$, $C = 0$ and $M' = \mathbf{U}'\left(\mathbf{U}\mathbf{U}'\right)^{-1}$, the necessary and sufficient conditions for independence are satisfied and we have established that $\hat{\Gamma}$ and $\hat{B}$ are independently distributed.

3.3. Sufficiency and completeness

We have $\left(\mathbf{V}-\Gamma\mathbf{U}\right)\left(\mathbf{V}-\Gamma\mathbf{U}\right)' = \left([\mathbf{V}-\widehat{\Gamma}\mathbf{U}]+[\widehat{\Gamma}-\Gamma]\mathbf{U}\right)\left([\mathbf{V}-\widehat{\Gamma}\mathbf{U}]+[\widehat{\Gamma}-\Gamma]\mathbf{U}\right)'$.
Then, by developing and using Eq.(11), we obtain:

$$\begin{aligned}\left(\mathbf{V}-\Gamma\mathbf{U}\right)\left(\mathbf{V}-\Gamma\mathbf{U}\right)' &= \left(\mathbf{V}-\widehat{\Gamma}\mathbf{U}\right)\left(\mathbf{V}-\widehat{\Gamma}\mathbf{U}\right)' + \left(\widehat{\Gamma}-\Gamma\right)\mathbf{U}\mathbf{U}'\left(\widehat{\Gamma}-\Gamma\right)' \\ &= (n-1)\widehat{B} + \left(\widehat{\Gamma}-\Gamma\right)\mathbf{U}\mathbf{U}'\left(\widehat{\Gamma}-\Gamma\right)'.\end{aligned}$$

The latter equation can be written as:

$$\begin{aligned}\mathbb{L}_{\mathbf{V}}(\Gamma, B) &= (2\pi)^{-(n-1)} \mid B \mid^{-\frac{n-1}{2}} \\ &\exp\left\{-\frac{1}{2} tr\left[B^{-1}\left((n-1)\widehat{B} + \left(\widehat{\Gamma}-\Gamma\right)\mathbf{U}\mathbf{U}'\left(\widehat{\Gamma}-\Gamma\right)'\right)\right]\right\}\end{aligned}$$

which means that $\left(\widehat{\Gamma}, \widehat{B}\right)$ is conjointly sufficient for (Γ, B)

The completeness of $\left(\widehat{\Gamma}, \widehat{B}\right)$ follows, by reasoning similar to that employed for the maximum likelihood estimators of the parameters of the multivariate normal distribution (see, for example, Anderson [2]).

Finally, as the estimators $\widehat{\Gamma}$ and $\dfrac{n-1}{n-3}\hat{B}$ are unbiased for Γ and B respectively, then we deduce that they are the UMVUE.

Acknowledgements

This work was supported partially by research project MTM 2005-09209, Ministerio de Educación y Ciencia, and project P06-FQM-02271, Junta de Andalucía, Spain.

References

[1] AIT-SAHALIA, Y. Maximum-likelihood estimation of discretely-sampled diffusion: A closed-form approximation approach. *Econometrica*, 70 (2002), 223–262.

[2] ANDERSON, T.W. *An introduction to multivariate statistical analysis.* Second Edition. Wiley. New York, 1984.

[3] ARNOLD, L. *Stochastic differential equations*. John Wiley and Sons, New York, 1973.

[4] BIBBY, B.M., SORENSEN, M. Martingale estimation functions for discretely observed diffusion processes.*Bernoulli*, 1(1/2) (1995), 17–39.

[5] FRANK, T.D. Multivariate Markov processes for stochastic systems with delays: application to the stochastic Gompertz model with delay. *Physical Review E*, 66 (1) (2002), 011914.

[6] GUTIÉRREZ, R., ANGULO, J.M., GONZÁLEZ, A., PÉREZ, R. Inference in lognormal multidimensional diffusion process with exogenous factors: application to modelling in economics. *Applied Stochastics Model and Data Analysis*, 7 (1991), 295–316.

[7] GUTIÉRREZ, R., GONZÁLEZ, A., TORRES, F. Estimation in multivariate lognormal diffusion process with exogenous factors. *Applied Statistics*, 4(1) (1997), 140–146.

[8] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A. Maximum likelihood estimation in multivariate lognormal diffusion process with a vector of exogenous factors. *Monografías del Seminario Matemático García de Galdeano*, 31 (2004), 337–346.

[9] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A. A bivariate stochastic Gompertz diffusion model: statistical aspects and application to the joint modelling of the Gross Domestic Product and CO_2 emissions in Spain. *Environmetrics*, 19 (2008), 643–658.

[10] GUTIÉRREZ, R., GUTIÉRREZ SÁNCHEZ, R., NAFIDI, A., MERBOUHA, A. A Stochastic diffusion model based on the Gamma density: Statistical inference. *Monografías del Seminario Matemático García de Galdeano*, 34 (2008), 117–125.

[11] GUTIÉRREZ, R., GUTIÉRREZ SÁNCHEZ, R., NAFIDI, A. The trend of the total stock of the private car-petrol in Spain: Stochastic modelling using a new Gamma diffusion process *Applied Energy*, 86 (2009), 18–24.

[12] KHATRI, C. G. Conditions for Wishartness and independence of second degrees polynomial in a normal vector. *The annals of mathematical Statistics*, 33 (3) (1962), 1002–1007.

[13] MAGNUS, J.R., AND NEUDECKER, H. *Matrix differential calculus with applications in statistics and economics*, New York: Wiley, 1988.

[14] PRAKASA-RAO BSL. *Statistical inference for diffusion type process*, New York: Ed. Arnold, London and Oxford University press, 1999.

[15] TONGHUI WANG. Versions of Cochran's theorem for general quadratic expressions in normal matrices. *Journal of Statistical planning and inference*, 58 (1997) 283–297.

Nafidi A.
Université Hassan 1er
Ecole Supérieure de Technologie- Berrechid
B.P: 218, Berrechid, Maroc
nafidi@estb.ac.ma

Gutiérrez R. and Gutiérrez-Sánchez R.
Department of Statistics and Operations Research
University of Granada, Facultad de Ciencias
Campus de Fuentenueva s/n, 18071
Granada, Spain
rgjaimez@ugr.es and
ramongs@ugr.es

Monografías Matemáticas García de Galdeano **36**, 89–97 (2010)

PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS IN DISCRETE DISTRIBUTIONS USING δ-RECORDS

Lina Maldonado

Abstract. We present a selection of results on the properties of Maximum likelihood estimators (MLE) developed in [7] for some discrete distributions (Geometric, Zeta and Poisson distribution) from the information provided by δ-records (cf. [5]). Maximization in the estimation is performed by numerical methods and the properties of the estimators have been studied using simulations. We consider the properties of the estimators as bias and mean squared error. In this paper we have shown that the incorporation of information from δ-records provides better estimates that when we use only available information from usual records or weak-records. Also, we observe that in the case of the light-tailed discrete distributions, a smaller $|\delta|$ is needed for better quality in the MLE than in other distributions.

Keywords: Maximum Likelihood Estimation, Weak Record, δ-records, Mean squared error, Geometric distribution, Zeta distribution, Poisson Distribution, Bias.

AMS classification: 60G70, 62F12

§1. Introduction

In the literature there have been many efforts to study records, increased in recent years. The motivations of the mathematicians in this study have been, among others, the large amount of data on them in various fields such as sports, hydrology, meteorology and industry -to name a few-, and the possibility of creating with these data, models that can predict the future record values.

Suppose observations come sequentially $X_1, X_2, \ldots$. As we observe then, we can also keep track on the maximum values reached by these observations, $M_1, M_2, \ldots$; several questions can be asked about it, for example, if we can calculate how often there is a change in this maximum. A change in the maximum value means that a value "record" has been observed. Another question which may be raised whether these values of records are given by some model. To answer these questions and some others, we need to see these values of records under terms of stochastic process.The first time that a statistical study introduced the notion of record was in 1952, with a basic model of records with observations *i.i.d.* made by Chandler [2].

Definition 1. Let $X_1, X_2, \ldots, X_k$ a sequence of *i.i.d.* random variables with common cumulative distribution *cdf* F. An observation X_j will be called an upper record (or simply a record), if it value exceeds all previous observations. Thus, X_j is a record if $X_j > X_i$ for every $i < j$.[1]

Weak records are obtained by replaced $>$ by $\geq$ in Def.1. This concept was introduced by [9] in the context of integer valued random variables, where the repetition of a record is also counted as record. More result on the behavior of records and weak records can be consulted in [1].
One concept derived from the definition of records is δ-records. Let $\delta \in \Re$, an observation X_j is called a δ-record if $X_j > X_i + \delta$, for $i < j$, that is, if it is greater than the previous maximun plus a (negative or positive) fixed quantity. For $\delta < 0$, every record is a δ-record, while for $\delta > 0$ this is not the case. In [4][6], the authors introduced a law of large numbers for the counting proccess of δ-records, in discrete and continuous distributions and a central limit theorem in discrete distributions. They obtained that the number of δ-records, with $\delta < 0$, grows at the same speed as the number of records (that is, as the logarithm of the number of observations) in heavy tailed distributions, proportional to the number of records in distributions with exponential tails and, faster in lighter tailed distributions. In [7] also studied, for $\delta < 0$, how to calculate the likelihood of a sample, consisting of n records and associated δ-records -both for continuous and discrete distributions-, and using that expression to calculate the *MLE* of parameters of several distributions (in particular, exponential distribution in continuous case and geometric distribution in discrete case).
The pioneers in considering the *MLE* of a parameter of a distribution from records were Samaniego and Whitaker [8], considering the parameter estimate σ of distribution $Exp(\sigma)$ using data $R_0^{'}, K_0, R_1^{'}, K_1, \ldots$, being $R_0^{'}, R_1^{'}, \ldots$ the successive minima, and $K_0, K_1, \ldots$ the number of trials needed to obtain new records. They also studied the nonparametric version of this problem. Some authors have studied the maximum likelihood estimation of parameters of discrete distributions, e.g. Doosparast and Ahmadi consider the estimation of the parameter θ of a geometric distribution based on record values, via *MLE* [3].
In the present study, we present a selection of results of the properties of Maximum likelihood estimators *MLE* developed in [7] for some discrete distributions (Geometric, Zeta and Poisson distribution) from the information provided by δ-records. The paper is organized as follows: in the next section, we introduce a short description of how δ-record statistics can be used to estimate the parameters of the parent distribution F, performed in [7]. In section 3, we compute the *MLE* of the parameters in some distributions. In section 4, we study the statistical properties of the *MLE* obtained in section 3.

§2. Pre-Implementation Results

In the rest of the paper δ will be a negative integer. The results of this section are contained in [7]. We consider nonnegative, integer-valued *i.i.d.* random variables $X_n, n \geq 1$ with common distribution F, such that $P[X_n = k] = p_k > 0$ for $k \in \mathbb{Z}_+$ and $n \geq 1$, with $p_m = 0$ if $m \leq -1$. The hazard rate r_k is defined by,

$$r_k = P[X_i = k | X_i \geq k] = \frac{P[X_i = k]}{P[X_i \geq k]}$$

The sample is obtained from a series of observations of nonnegative, *i.i.d.* random variables. Instead of collecting only records values, we have δ-records values (when a record value is observed, we also collect the observations at a distance less than $|\delta|$ of record R_i until the

next record R_{i+1} appear), and so until there is an observation higher than a predetermined amount n. Is important remark that, the δ-records associated to a record need not be ordered.

We compute the number of times that the integer j is equal to the previous maximum - k, with $j + k \leq n$, denoted by ξ_j^k, with $0 \leq k \leq |\delta|$ and $0 \leq j \leq n - k$.

$$\xi_j^k = \sum_{m:M_m \leq n} \mathbf{1}_{\{X_m = j, M_m = j+k\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

From this, the likelihood of the sample is given by,

$$P\left\{\xi_0^0 = i_0^0, \ldots, \xi_n^0 = i_n^0, \xi_0^1 = i_0^1, \ldots, \xi_{n-1}^1 = i_{n-1}^1, \ldots, \xi_0^k = i_0^k, \ldots, \xi_{n-k}^k = i_{n-k}^k\right\} \propto \prod_{j=0}^{n} r_j^{s_j^k} \left(1 - r_j\right)^{1+\sum_{l=1}^{k} s_{j+l}^{k-l}} \quad (1)$$

with $k = -\delta$, r_j is the hazard rate and $s_j^i = \sum_{m=0}^{i} i_j^m$ the number of times j has been a δ-record at a distance $\leq i$ of the previous maximum.

§3. Statistical inference on the model

3.1. Maximum likelihood estimators of parameters of discrete distributions

From the previous results on the likelihood of the sample (1), we are going to calculate maximum likelihood estimates of three distributions (Geometric, similar to the Poisson and similar to Zeta). These distributions are studied for their characteristics: a similar to the Poisson distribution is a light-tailed distribution with $r_k \to 1$, the geometric distribution with $r_k = r$ (r constant) and the similar to Zeta distribution, is a heavy-tail distribution with $r_k \to 0$.

3.1.1. MLE $\hat{p}$ in a geometric distribution

Expression (1) can be used to find the *MLE* $\hat{p}$ of the parameter of the geometric distribution, $Geom(p)$. Since $r_k = p$, with $k = 0, 1, \ldots$, we have

$$\mathcal{L}(p) \propto p^{\sum_{j=0}^{n} s_j^k} \left(1 - p\right)^{(n+1)+\sum_{j=0}^{n}\sum_{l=1}^{k} s_{j+l}^{k-l}} \quad (2)$$

From this, the log-likelihood function of equation (2) becomes

$$\sum_{j=0}^{n} s_j^k \log\left(p\right) + \left[(n+1) + \sum_{j=0}^{n}\sum_{l=1}^{k} s_{j+l}^{k-l}\right] \log\left(1 - p\right)$$

and finally, we have that *MLE* of p is given by

$$\hat{p} = \frac{\sum_{j=0}^{n} s_j^k}{(n+1) + \sum_{j=0}^{n} \sum_{l=1}^{k} s_{j+l}^{k-l}} \tag{3}$$

3.1.2. MLE $\hat{\lambda}$ in a similar to poisson distribution

Similarly to the previous subsection, we take the equation (1) for maximum likelihood estimation of a discrete variable λ (similar to the Poisson) with a failure rate $r_k = 1 - \frac{\lambda}{k+a}$, with $k = 0, 1, \ldots$ and a known. Note that r_k in the Poisson distribution has $r_k \to 1$, so this distribution is, in fact, very close to Poisson. Performing a calculation similar to equation (2), we obtain,

$$\log(\mathcal{L}) = \sum_{j=0}^{n} s_j^k \log\left(1 - \frac{\lambda}{j+a}\right) + \sum_{j=0}^{n} \left(1 + \sum_{l=1}^{k} s_{j+l}^{k-l}\right) \log\left(\frac{\lambda}{j+a}\right)$$

Thus, with $a_j = s_j^k$ and $b_j = \left(1 + \sum_{l=1}^{k} s_{j+l}^{k-l}\right)$, the value λ that maximizes the likelihood is given by

$$\sum_{j=0}^{n} \frac{\frac{a_j}{j+a}}{\left(1 - \frac{\lambda}{j+a}\right)} + \frac{\sum_{j=0}^{n} b_j}{\lambda} = 0 \tag{4}$$

Equation (4), can be solved numerical calculation, to yield the maximum likelihood estimator of the parameter λ. It is important to note that, in similar studies using record values maximization is done using numerical analysis.

3.1.3. MLE $\hat{a}$ in a similar to zeta distribution

Now, to calculate the *MLE* for a distribution similar to Zeta distribution. The $Zeta(a)$ is given by $p_k = \frac{(k+1)^{-a}}{\zeta(a)}$ for $k \in \mathbb{Z}_+$, $a > 1$ with $\zeta(a) = \sum_{j=0}^{\infty}(j+1)^{-a}$ and $r_k = \frac{(k+1)^{-a}}{\sum_{j=k}^{\infty}(j+1)^{-a}}$. We consider a distribution similar to zeta with a failure rate $r_k = \frac{a}{j+b}$, $j = 0, 1, \ldots$, with $a > 1$ unknow and b known. Then, considering the likelihood function (1) we have

$$\mathcal{L}(a) \propto \prod_{j=0}^{n} \left(\frac{a}{j+b}\right)^{s_j^k} \left(1 - \left(\frac{a}{j+b}\right)\right)^{1+\sum_{l=1}^{k} s_{j+l}^{k-l}}$$

then,

$$\log(\mathcal{L}) = \sum_{j=0}^{n} a_j \ln(a) + \sum_{j=0}^{n} b_j \ln\left(1 - \frac{a}{j+b}\right) \tag{5}$$

where $a_j = s_j^k$ and $b_j = 1 + \sum_{l=1}^{k} s_{j+l}^{k-l}$. As in 3.1.2, the equation $\log(\mathcal{L}) = 0$ can be solved numerically.

§4. Results and Conclusions

4.1. Properties of likelihood estimators

The properties of the *MLE* are studied via computer simulations, in the three distributions described above, and worked with a number of 10.000 replicates varying the maximum n which fix the stopping value for the simulation run. We start with the results of the estimation of the similar to Poisson distribution parameter $(\lambda = 3)$. In Fig.1 we present the boxplots corresponding to the estimates using information obtained from maximum - δ, with $|\delta| = 0 \ldots 5$, where $|\delta| = 0$ correspond to records values and $|\delta| = 1$ to weak records, and $n = 12$. In the graph, we observe much greater variability in the boxplot corresponding to the estimate made only with records and weak records than in the case $|\delta| > 1$, that is, the estimations are better when $|\delta|$ grows. Together with the decrease in variability, more simetric is observed in the case $|\delta| = 5$ than in the case $|\delta| = 0$. The decrease of variability for $|\delta| = 5$ is almost 70% with respect to use of records.

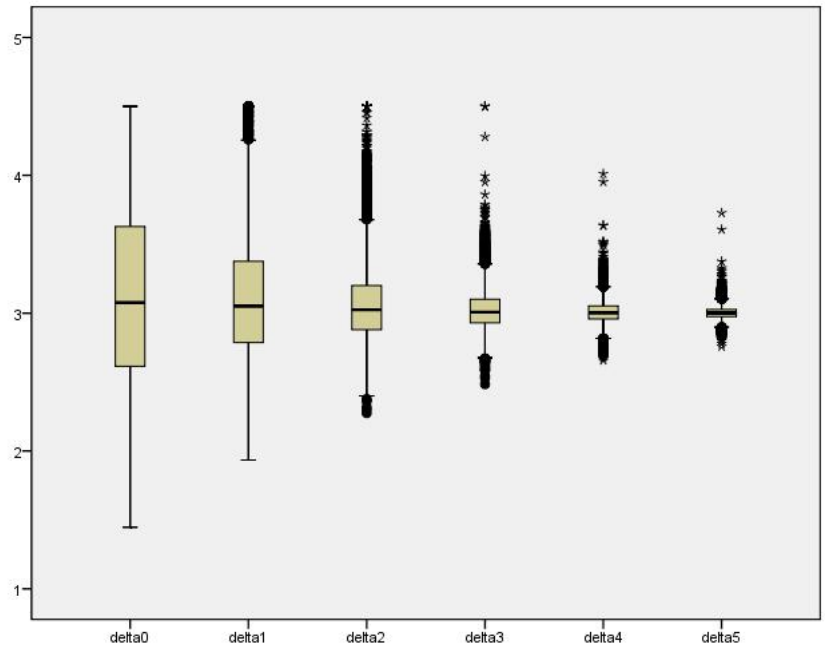


Figure 1: Boxplot-10000 simulations of $\lambda = 3$ of a similar to Poisson Distribution, with max=10

In Fig.2, the properties of the *MLE* can be evaluated by studying of the values of the estimations and how acumulate around the real value of $\lambda = 3$.
Studying the Normal QQ-Plot, these observations are not sufficiently close to the line and therefore, we can not establish that the *MLE* has a normal distribution (looks heavier tails than normal), see Fig.3. Moreover, this is confirmed by K-S test. It is worth mentioning here

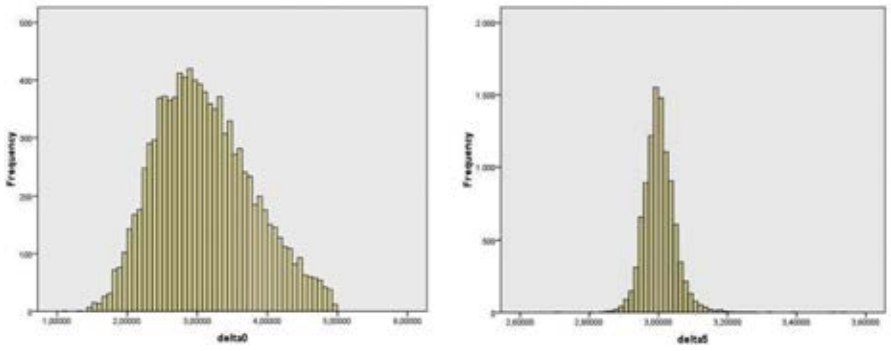


Figure 2: Histograms - 10000 simulations of $\lambda = 3$ of a similar to Poisson Distribution, with max=10 ($\delta = 0$ and $\delta = 5$, respectively)

that, in other distributions, the same non-normal behavior was observed for the estimators.

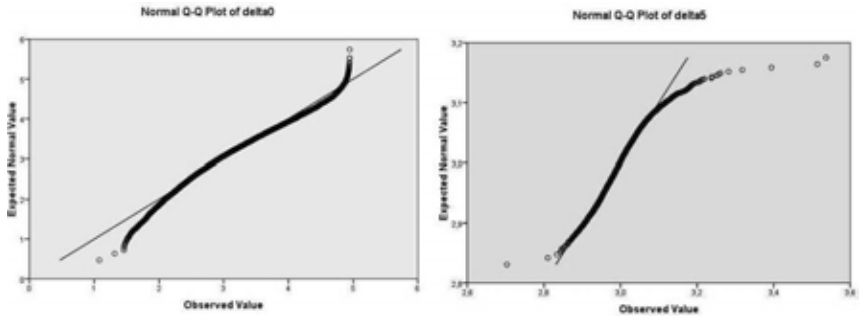


Figure 3: QQplot-10000 simulations of $\lambda = 3$ of a similar to Poisson Distribution, with max=10 ($\delta = 0$ and $\delta = 5$, respectively)

Now, in Table 1, we can see the *MSE* and bias of parameter, showing a significant drop in the *MSE* by incorporating the information from δ-records. Also, we can say that the poisson distribution achieves the biggest change in the order of the *MSE* when comparing the estimates obtained by the records values against the *MSE* with information of $\delta = -5$ of all three distributions studied. This order of variation is about $\frac{MSE(\delta=0)}{MSE(\delta=-5)} = 238.92$ in the case of simulating until maximum 10. In the case of a maximum equal to 12 this order is doubled Next, we present the results of the estimation of the geometric distribution with parameter $p = 0.5$, simulated up to a maximum of 20 (see Fig.4). As in the previous case, the graph shows greater variability in the boxplot corresponding to the estimate made only with records values. In this case, there is a decrease of MSE around 67% (the variation order for maximum until 10 is $\frac{MSE(\delta=0)}{MSE(\delta=-5)} = 7.05$, for $Max = 15$ this rate is doubled and $Max = 20$ is tripled). Also, as we see in Table 2, the bias reduction was around 68 % with respect to the estimation based on record values only.

Simulations	Maximum	Measures	δ-Records	
			0	-5
10000	**10**	**MSE**	0,48296	0,00202
		Bias	-0,10480	-0,00313
	12	**MSE**	0,47125	0,00093
		Bias	-0,13403	-0,00183

Table 1: *MSE* and Bias of $\lambda = 3$ of a similar to Poisson Distribution

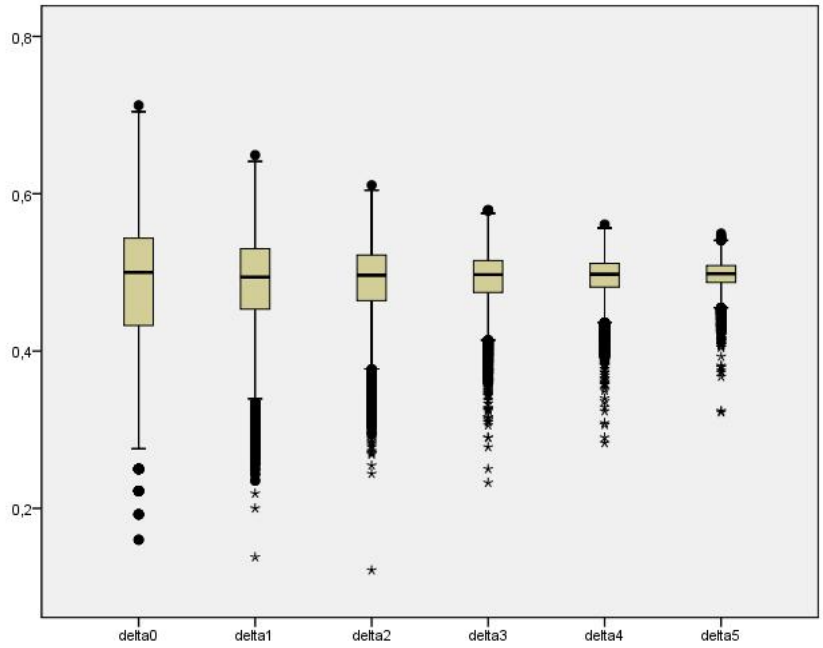


Figure 4: Boxplot - $p = 0.5$ of a Geometric Distribution, with max=20

Simulations	Maximum	Measures	δ-Records	
			0	-5
10000	**10**	**MSE**	0,01260	0,00178
		Bias	-0,02293	-0,01050
	20	**MSE**	0,00636	0,00033
		Bias	-0,01098	-0,00344

Table 2: *MSE* and Bias of $p = 0.5$ of a Geometric Distribution

In the zeta distribution, the estimates made by incorporating the information from δ-records (especially in the case of $\delta = -5$), produced better estimates as in previous distributions;

however, the differences between them are not so marked. In this case, there is the presence of many outliers in the estimates for the different δ-records, as seen in Fig.5. The *MLE* with $\delta = -5$ have high *MSE* (unlike previous distributions) and the negative bias is also high.
We consider now the boxplots in the estimates of the parameter $a = 4$ of zeta distribution. As we see, there is a substantial improvement of the estimates (see Fig. 5). This is confirmed in Table 3, where we see that the variation rate is $\frac{MSE(\delta=0)}{MSE(\delta=-5)} = 6.74$.

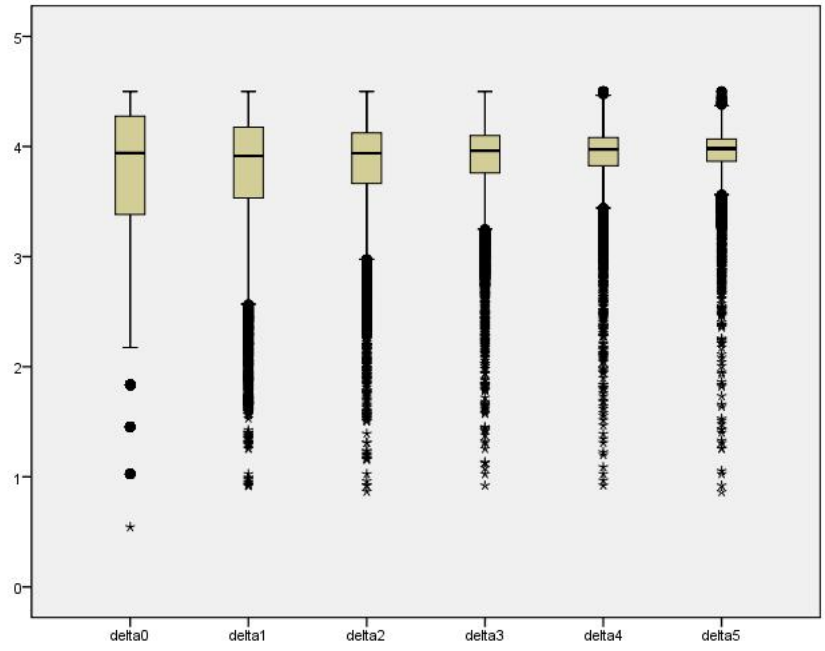


Figure 5: Boxplots - $a = 4$ of a similar to Zeta Distribution, with max=20

Simulations	Maximum	Measures	δ-Records	
			0	-5
10000	**10**	**MSE**	0,59950	0,09513
		Bias	-0,28096	-0,07759
	20	**MSE**	0,50051	0,07416
		Bias	-0,25027	-0,06996

Table 3: *MSE* and Bias of $a = 4$ of a similar to Zeta Distribution

4.2. Conclusions

In this paper, we have shown that the incorporation of information from δ-records provides better estimates than when we use only available information from usual records or weak-

records. We show that while $|\delta|$ grows, the estimates improve. Also, we noted that for light tailed distributions, a delta smaller is required to obtain good results. In all cases, the distribution of the *MLE* using information provided by δ-records does not appear to be normal. Finally, it is remarkable that the maximum likelihood estimates can be difficult to calculate explicity, so it is necessary to use numerical methods for maximization.

§5. Future Work

As future work we consider the likelihood function by fixing the number of records in the sample and propose confidence intervals for the Maximum Likelihood Estimators of the parameters of the distributions. Other proposals, such as the estimation of the hazard rate using non-parametric estimation or the use of arguments similar to those developed in this work, for F continuous, can also be considered.

Acknowledgements

The autor wants to acknowledge the financial support granted from Fundación Gran Mariscal de Ayacucho *(Fundayacucho)* scholarship.

References

[1] ARNOLD, B., BALAKRISHNAN, N. AND NAGARAJA, H. Records. *John Wiley & Sons*, New York, (1998).

[2] CHANDLER, K.N. The distribution and frequency of record values. *J. Roy. Statist. Soc. Ser.* **14**, (1952), 220–228.

[3] DOOSTPARAST, M. AND AHMADI, J. Statistical analysis for geometric distribution based on records, *Computers and Mathematics with Applications*, **52**, (2006), 905–916.

[4] GOUET, R., LÓPEZ F.J. AND SANZ G. Central limit theorems for the number of records in discrete models, *Adv. Appl. Prob*, **37**, (2005), 781–800.

[5] GOUET, R., LÓPEZ, F.J., AND SANZ, G. Asymptotic normality for the counting process of weak records and δ-records in discrete models, *Bernoulli* **13** (3), (2007), 754–781.

[6] GOUET, R., LÓPEZ, F.J., AND SANZ, G. Laws of Large numbers for the number of weak record, *Statistics & Probability Letters* **78**, (2008), 2010–2017.

[7] GOUET, R., LÓPEZ, F.J., AND SANZ, G. On the structure of δ-records values, (2011). Preprint.

[8] SAMANIEGO, F.J. AND WHITAKER, L.R. On estimating population characteristics from record-breaking observations. I: Parametric results, *Naval Research Logistics Quarterly*, **33**(3), (1986), 531–543.

[9] VERVAAT, W. Limit theorems for records from discrete distributions, *Sthocastic Processes Appl*, **1**, (1973), 317–334.

Lina Patricia Maldonado Guaje
Dpto. Métodos Estadísticos
Universidad de Zaragoza
lina.maldonado@unizar.es

Monografías Matemáticas García de Galdeano **36**, 99–108 (2010)

RE-IMPLEMENTING NSGA-II AND SPEA2 USING PARETO BASED OPERATORS

P.M. Mateo, I. Alberto

Abstract. In [2, 3] a mutation and a crossover operators, both based on Pareto ranking, were presented. Individually, they showed very good performance when compared with other usual operators. In this work we re-implement two well-known Evolutionary Algorithms: NSGA-II and SPEA2, using our operators and three different Differential Evolution variation operators. Then, the results obtained with the original versions of the algorithms and those using Differential Evolution are compared with those reached when introducing our operators in the algorithms. This comparison allows us to show that our operators behave better, in general, than the others.

Keywords: Multiobjective Decision Making, Metaheuristics, Evolutionary Algorithms

AMS classification: 90C29, 90C59, 68T20

§1. Introduction

Evolutionary Algorithms, EAs, can be considered the most adequate methods for solving complex Multiobjective Optimisation Problems (MOOPs). It was in [14], the PhD of the author, where the Vector Evaluated Genetic Algorithm (VEGA) was presented. It constituted the first actual implementation of what has been called a Multiobjective Evolutionary Algorithm (MOEA). Since then, several different MOEAs have appeared. An extensive review on this matter can be obtained, for instance, from [5] and [6].

NSGA-II by [7] and SPEA2 by [18] are two of the most competitive MOEAs available in the literature and they are often used as a reference to determine the performance of new approaches. Differential Evolution, DE, proposed by [15], is a relatively new heuristic technique initially designed for optimizing single-objective continuous problems. However, DE has also been extended to the multiobjective case as for example in [1, 13, 16, 12].

In this work we present a re-implementation of the algorithms NSGA-II and SPEA2 using the operators presented in [2, 3], as well as three different DE variation operators. The results obtained with the original versions and those using DE are compared with the results reached with our operators. This comparison allows us to show that our operators behave better, in general, that the other ones in the proposed test problems.

§2. Multiobjective Optimisation and Evolutionary Algorithms

The aim of Multiobjective Optimisation is to optimise a set of objective functions which, in general, may be of a conflicting nature. Hence, the term “optimise” means to find a solution

satisfying the constraints, which would give reasonable values of all objective functions to the decision maker. Formally, Multiobjective Optimisation Problems can be defined as:

$$\min \vec{f}(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}) \quad \text{s.t. } \vec{x} = (x_1, \ldots, x_n) \in D \subset \mathbb{R}^n$$

Contrary to single objective optimisation, in multiobjective optimisation it is usually impossible to find one optimal solution. Instead, algorithms for optimising multiobjective problems try to find a family of points known as the *Pareto optimal set*. These points verify that there is no different feasible solution which strictly improves one component of the objective function vector without worsening at least one of the remaining ones. The two next definitions describe mathematically these ideas.

Definition 1. If given a solution $\vec{y}$, there exists another solution $\vec{x}$ such that $\forall j = 1, \ldots, m$ $f_j(\vec{x}) \leq f_j(\vec{y})$ and $\exists j \in \{1, \ldots, m\}$ such that $f_j(\vec{x}) < f_j(\vec{y})$, then we will say that solution $\vec{x}$ *dominates* solution $\vec{y}$ (denoted by $\vec{x} \prec \vec{y}$).

Definition 2. A solution $\vec{x} \in D$ is said to be *Pareto optimal* or *efficient* if and only if $\nexists \vec{y} \in D$ such that $\vec{y} \prec \vec{x}$. The set of Pareto optimal solutions $\vec{x} \in D$ is called *Pareto optimal set* and it will be denoted with P^{true}. The image of P^{true} in the objective function space is called *Pareto front* and it will be denoted with PF^{true}.

A clear description of the Strength Pareto Evolutionary Algorithm, SPEA2, and the Non-dominated Sorting Genetic Algorithm, NSGA-II, (Evolutions of the original versions SPEA and NSGA, respectively), can be found, for instance, in [5]. They propose techniques for approximating the Pareto-optimal set for MOOP's and they have shown very good performance in comparison to other MOEAs. For a detailed description of both algorithms the readers are referred to [18] and [7].

§3. Operators based on Pareto rankings

In the experiments carried out we compare different versions of the algorithms mentioned above. As it will be explained later in Section 4, in those algorithms we have substituted the original mutation and crossover operators by the ones proposed in [2, 3], respectively. Both operators are devoted to both exploitation and exploration depending on the quality of the solution. The main idea of the operators is to use good parents (efficient ones) to improve the quality of the offspring (exploitation) and to use not so good parents (non-efficient ones) to explore the whole space (exploration). The outlines of these Pareto Based Operators, PBO, are shown in the following paragraphs.

The rank based mutation operator, $RBMO$, has a doubly tuning behaviour. The amplitude of the mutation depends on both, the iteration and an appropriate Pareto ranking of the individual. In order to establish the rank of the individuals we have used the ranking proposed by [10], but any other could be used. The novelty of the operator lies in the fact of using a Pareto ranking (and therefore, the quality of the solution) to control the maximum amplitude: the smaller the rank, the smaller the maximum amplitude.

Let $\vec{x}^{(h,t)} \in P^t$ be a solution in the iteration t of the algorithm and let k be the component that is going to be mutated. It operates in the following way:

$$x_k^{(h,t+1)} = x_k^{(h,t)} + 2\delta_k(x_k^{(U)} - x_k^{(L)})(u - 0.5)$$

where u is a random number in $[0,1]$, $\delta^{(t,r)}$ defines the amplitude, and it depends on both t and r according to the expression:

$$\delta^{(t,r)} = \delta_k(t,r) = \begin{cases} 2\epsilon(t) + 0.0001[1-\epsilon(t)], & \text{if } r_{\max} = 1 \\ \epsilon(t) + [1-\epsilon(t)]\lambda(r), & \text{if } r_{\max} \geq 2 \end{cases}$$

where

$$\epsilon(t) = 0.25\left(\frac{t_{\max} - t}{t_{\max}}\right)^2 \text{ and } \lambda(r) = \left(\frac{1-e^{-\beta(r-1)}}{1-e^{-\beta(r_{\max}-1)}}\right),$$

being $t_{\max}$ the maximum number of iterations allowed and $r_{\max}$ representing the maximal rank of the individuals of the population in the current iteration. The parameter β in $\lambda(r)$ is a shape parameter which allows us to establish the trade-off between efficient and non-efficient solutions. The value $\beta = -0.1$ has been taken as in [2]. The implementation of our mutation operator also considers the incorporation of a modification of the probability of mutation, which is $2p$ for the individuals with rank greater than $[r_{\max}/2]$, being p for the remaining ones.

The Pareto based crossover operator, PBC, is a uniform crossover operator which has a different maximum range of variation depending on the quality of the solutions. This operator takes into consideration whether the solutions to be recombined are efficient or not and whether the solutions to be recombined are close to each other or not.

Let $\mathcal{P}^t$ be the current population in iteration t, POS^t its Pareto optimal set and $p_t = |POS^t|$. In order to decide whether or not the parents are close to each other, a fact that will be used in the design of the operator, we use the following process:

Step 1: Let $dist(\vec{x},\vec{y})$ be a distance measure between $\vec{x}$ and $\vec{y}$. Calculate $d^{(\text{eff},t)}$ the mean distance between all pairs of solutions in POS^t and $\sigma^{(\text{eff},t)}$ the standard deviation. If there is only one efficient solution, we set $d^{(\text{eff},t)} = 0$.

Step 2: Let f^{Neff} and $f^{(\text{eff},t)}$ be the amplitude factors associated to the non-efficient and efficient solutions, respectively, calculated according to

$$f^{\text{Neff}} = 1 \text{ and } f^{(\text{eff},t)} = 1 - \lfloor 10 \cdot \frac{t}{t_{max}} \rfloor \cdot 0.11 \quad (\lfloor\cdot\rfloor \equiv \text{ floor function})$$

After the above establishments, let $\vec{x}^{(1,t)}$ and $\vec{x}^{(2,t)}$ be the parent solutions selected for crossover. At that moment, and depending on the distance between $\vec{x}^{(1,t)}$ and $\vec{x}^{(2,t)}$ and on the values of $d^{(\text{eff},t)}$ and $\sigma^{(\text{eff},t)}$, the factors $f^{(\text{eff},t)}$ and f^{Neff} are both multiplied by one of the following quantities:

(i)	1	if $d^{(\text{eff},t)} = 0$ or $\sigma^{(\text{eff},t)} = 0$
(ii)	$1 - \text{sign}(d) \cdot \frac{1}{3}$	if $\lvert d\rvert > 3$
(iii)	$(1 - d \cdot 0.11)$	if $-3 \leq d \leq 3$

where $d = (dist(\vec{x}^{(1,t)},\vec{x}^{(2,t)}) - d^{(\text{eff},t)})/\sigma^{(\text{eff},t)}$. Finally, a last correction in $f^{(\text{eff},t)}$ is done: $f^{(\text{eff},t)} = \min\{1, f^{(\text{eff},t)}\}$.

After these calculations we denote $\Delta(k) = \frac{3}{4}|x_k^{(1,t)} - x_k^{(2,t)}|$. In order to carry out the crossover, three cases are considered:

- **Case 1**: Neither of the parents is efficient. In this situation, the new components $x_k^{(1,t+1)}$ and $x_k^{(2,t+1)}$, will be generated taking two random values in the interval

$$\left[\frac{x_k^{(1,t)} + x_k^{(2,t)}}{2} - \frac{4}{3}\,\Delta(k)\; f^{\text{Neff}}, \frac{x_k^{(1,t)} + x_k^{(2,t)}}{2} + \frac{4}{3}\,\Delta(k)\; f^{\text{Neff}}\right]$$

- **Case 2**: Both parents are efficient solutions. Then, two values, $x_k^{(h,t+1)}$, $h = 1, 2$, for the new solutions are randomly created in the intervals

$$\left[x_k^h - \Delta(k)\; f^{(\text{eff},t)}, x_k^h + \Delta(k)\; f^{(\text{eff},t)}\right],\ h = 1, 2,$$

  where, when we are obtaining $x_k^{(h,t+1)}$, x_k^h is equal to $x_k^{(h,t)}$ with probability 0.75 or equal to $x_k^{(3-h,t)}$ with probability 0.25, $h = 1, 2$.
- **Case 3**: Only one of the parents is efficient. Assuming that $\vec{x}^{(1,t)}$ is the efficient solution (the other case is analogous), then one value is obtained in each of the intervals

$$\left[x_k^1 - \Delta(k)\; f^{(\text{eff},t)}, x_k^1 + \Delta(k)\; f^{(\text{eff},t)}\right],\ \left[x_k^2 - \Delta(k)\; f^{\text{Neff}}, x_k^2 + \Delta(k)\; f^{\text{Neff}}\right]$$

  where, when we are obtaining $x_k^{(h,t+1)}$, x_k^h is equal to $x_k^{(h,t)}$ with probability 0.75 or equal to $x_k^{(3-h,t)}$ with probability 0.25, $h = 1, 2$.

In both operators, $RBMO$ and PBC, the solutions are randomly generated in the corresponding intervals using a uniform distribution. Besides this, if the solution obtained is not feasible it is re-sampled until it fits the bounds.

§4. Elements of the computational experiment

In this section we present the elements carried out in order to accomplish the experiment of Section 5. Regarding to the elaborated codes and the executions, we indicate that all programmes were coded in C language and compiled with a GNU GCC compiler. All the codes used in the paper were written by the authors except: The code for the hypervolume calculation that was obtained from [9], and all the codes for the function evaluations that were extracted from the codes provided by [4].

4.1. The test problems

The use of a set of test problems helps to guarantee that the proposed algorithms will confront efficient solution spaces of different characteristics. We have used the extended and rotated and/or shifted version of some of the well-known set of problems ZDT [17] and DTLZ [8] proposed by [11]. A short description of their characteristics is shown in Table 1. In all of these test problems, the Pareto optimal set is known and it is provided by [4].

Table 1: Properties of the extended and rotated/shifted versions of the corresponding original ones. S: Separable; NS: Nonseparable; U: Unimodal; M: Multimodal.

Test problem	Objective functions	No. of variables	Total no. of variables	Separability and modality	Geometry
1. S-ZDT1	f_1 / f_2	1 / >1	30	S, U / S, U	convex
2. S-ZDT2	f_1 / f_2	1 / >1	30	S, U / S, U	concave
3. S-ZDT4	f_1 / f_2	1 / >1	30	S, U / S, M	convex
4. R-ZDT4	$f_{1:2}$	>1	10	NS, M	convex
5. S-ZDT6	f_1 / f_2	1 / >1	30	S, M / S, M	concave
6. S-DTLZ2	$f_{1:3}$	>1	30	S, U	concave
7. R-DTLZ2	$f_{1:3}$	>1	10	NS, M	concave
8. S-DTLZ3	$f_{1:3}$	>1	30	S, M	concave

4.2. Variation operators

The crossover and mutation operators that NSGA-II and SPEA2 implement are the Simulated Binary Crossover and the Polynomial Mutation Operator. These operators are common in the literature. The readers can consult the book by [6] to obtain their descriptions.
On the other hand, Differential Evolution (DE) is a type of EA proposed by [15] whose main difference lies in the way in which the variation operator is implemented, it uses weighted differences between solutions to perturb the population instead of being based in probability density functions. The first authors that apply DE to MOOPs were [1] in the so-called Pareto Differential Evolution algorithm. There are several DE algorithms for multiobjective optimization in the literature, we have selected three of them and we have extracted their variation operators with the aim of incorporating them into NSGA-II and SPEA2:

DEMO [13] The variation operator extracted from this paper works as follows: Let $\vec{x}^{(h,t)}$ be the individual to be mutated, then the new individual is built according to:

$$\vec{x}^{(h,t+1)} = \vec{x}^{(h,t)} + F \cdot (\vec{x}^{(r_1,t)} - \vec{x}^{(r_2,t)}) \tag{1}$$

where $\vec{x}^{(r_j,t)}, j = 1, 2$, are randomly selected with $r_1 \neq r_2 \neq h$, and F is a parameter supplied by the user, we use $F = 0.3$.

MODE [16] In this work the definition of the mutation operator depends on whether the solution $\vec{x}^{(h,t)}$ to be mutated is efficient or not.

$$\text{Efficient} \quad \vec{x}^{(h,t+1)} = \vec{x}^{(h,t)} + F \cdot \Big((\vec{x}^{(r_1,t)} - \vec{x}^{(r_2,t)}) + (\vec{x}^{(r_3,t)} - \vec{x}^{(r_4,t)})\Big)$$

$$\text{No Efficient} \quad \vec{x}^{(h,t+1)} = \gamma \cdot \vec{x}^{(h,t)}_{best} + (1-\gamma) \cdot \vec{x}^{(h,t)} + F \cdot \big((\vec{x}^{(r_1,t)} - \vec{x}^{(r_2,t)}) + (\vec{x}^{(r_3,t)} - \vec{x}^{(r_4,t)})\big).$$

Where $r_1 \neq r_2, r_3 \neq r_4$, $F = 0.5$, $\gamma = 0.7$ and $\vec{x}^{(h,t)}_{best}$ is a randomly selected individual from the set of non-dominated individuals of the population $\mathcal{P}^t$ that dominate $\vec{x}^{(h,t)}$.

NSDE [12] The mutation operator works as follows:

$$\vec{x}^{(h,t+1)} = \vec{x}^{(h,t)} + K \cdot (\vec{x}^{(r_3,t)} - \vec{x}^{(h,t)}) + F \cdot (\vec{x}^{(r_1,t)} - \vec{x}^{(r_2,t)}).$$

In the expression we take $K = 0.4$ and $F = 0.8$, furthermore, $r_i \neq r_j,\ \forall i \neq j$.

After the mutation, in the three cases a crossover operator is applied. The crossover operator acts in the following way:

$$x_i^{(h,t+1)} = \begin{cases} x_i^{(h,t+1)} & \text{if } rand(i) \leq CR \text{ or } i = rnbr(n) \\ x_i^{(h,t)} & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, n$$

where $rand(i)$ is a random value in $[0, 1]$, $CR \in [0, 1]$ is the crossover probability and $rnbr(n)$ is a randomly chosen index component in $\{1, 2, \ldots, n\}$.
In all the cases, the values of F, K and CR that we have used are those proposed by the authors in their respective papers.

4.3. Measures for comparing the populations

In order to compare the different implementations of the algorithms NSGA-II and SPEA2, we have considered three measures: Hypervolume difference to a reference set HD [4, 11], generational distance GD and set coverage C. The book by [5] can be consulted for a detailed description of these two last indicators.
Given P^* the set of non-dominated solutions resulting from the execution of an optimisation algorithm and PF^* its image in the objective space, HD calculates the difference between the hypervolume of PF^{true} and the hypervolume of PF^*, this measure takes into consideration the spread as well as the proximity of PF^* respect to PF^{true}. GD reports how far, on average, PF^* is from PF^{true}. For both measures the smaller the value the better the obtained population is. The real Pareto fronts, PF^{true}, are provided by [4]. The third measure, C, is a binary measure, given P_1^* and P_2^* the non-dominated sets resulting from the execution of two different algorithms, C calculates the proportion of solutions in set P_2^* which are weakly dominated by solutions in set P_1^* and it is denoted as $C(P_1^*, P_2^*)$; and reciprocally $C(P_2^*, P_1^*)$. If $C(P_1^*, P_2^*)$ is greater than $C(P_2^*, P_1^*)$, then P_1^* will be better than P_2^* and vice versa.

§5. Computational experiment and results

First of all, we have to point out that the aim of this work is not to compare the algorithms NSGA-II and SPEA2 but to show information about how the mutation and crossover operators we presented in [2, 3] improve the behaviour of these classical algorithms. With all the previous information, the implemented algorithms are shown in Table 2.

Table 2: List of variations implemented in the algorithms NSGA-II and SPEA2.

Name	Alteration in NSGA-II and SPEA2
(NSGAII/SPEA2)-O	None: Original versions of the algorithms
(NSGAII/SPEA2)-PBO	$RBMO$ and PBC variation operators
(NSGAII/SPEA2)-DE1	DE variation operator of $DEMO$
(NSGAII/SPEA2)-DE2	DE variation operator of $MODE$
(NSGAII/SPEA2)-DE3	DE variation operator of $NSDE$

For carrying out the experiments showed in this section, we have considered a usual fixed population size equal to 100 individuals as in [18] and [7]. The values for the probabilities of mutation and crossover are, respectively, equal to $1/n$ and 0.9 for algorithms NSGA-II and SPEA2, as their authors use. For those algorithms that make use of the Pareto based operators, these values are equal to $1/n$ or $2/n$ according to the description given in the mutation operator, and for the crossover operator the probability is equal to 0.9. For the algorithms that use DE operators, the values of these probabilities are those that their authors proposed in their papers and commented on in subsection 4.2.
The comparison of the implementations of the algorithms of Table 2 is accomplished taking into account three different values for the number of iterations executed: 100, 500 and 1000. For each problem, 50 initial populations are generated. For each number of iterations and after applying each implementation we obtain 50 final populations. In each of these populations, we get the set of efficient solutions, P^*, and we calculate the measures of Subsection 4.3. For calculating HD, for each problem and number of iterations, the reference point is placed by obtaining the worst objective function value among all the P^* populations obtained. That is to say, for each problem the same reference point is used for all the P^* once the number of iterations has been fixed.
Since only the mean values themselves are not enough to ensure whether an implementation is better for a given measure than other or not, we have considered the carrying out of a statistical test to determine this fact. To perform the statistical hypothesis tests for comparing the implementations, the data are paired in the following way: For each number of iterations and initial population, the measure obtained in the resulting population using one implementation is paired with the measure obtained in the resulting population using another implementation.
After these preliminaries, we compare the implementations (NSGAII/SPEA2)-PBO with the implementations (NSGAII/SPEA2)-(O/DE1/DE2/DE3) of Table 2 in order to decide whether or not the ones we propose are better than the others attending to the measures of subsection 4.3. Table 3 shows the results of the statistical test accomplished, t-test for the paired difference of measures HD and GD, and C(non-PBO imp., PBO imp.)$-C$(PBO imp., non-PBO imp.) for coverage. A '$+$' sign represents that the hypothesis "PBO implementation is better than non-PBO implementation" cannot be rejected, an '$=$' sign represents that the hypothesis "PBO implementation is equivalent to the non-PBO implementation" cannot be rejected, and a '$-$' sign appears otherwise, representing that PBO implementation is not equal nor better than the non-PBO implementation. Then, in this tables, the fewer '$-$' signs appear, the better the PBO implementation is.
Before starting the analysis of the results obtained, note that, since there are three measures for judging the behaviour of the algorithms to be compared, the decision of which algorithm behaves better is, itself, a multiobjective problem. Then, we have decided to make the following compromise decision: For commenting on Table 3 we have organized the results in such a way that if the difference between the number of '$+$' and '$-$' signs in a triplet corresponding to HD, GD and C is positive, we will say that PBO implementation "performs better than" non-PBO implementation; if that difference is negative, we will say that non-PBO implementation "performs better than" PBO implementation; PBO and non-PBO implementations performing in an equivalent way otherwise.

S-ZDT1 When 100 iterations are considered, only the implementation based on DE1 of both

algorithms presents equivalent behaviour to PBO implementation. The remaining ones, those based on DE1 with 500 and 1000 iterations and on O, DE2, DE3 with 100, 500 and 1000 iterations, show a worse behaviour than the ones based on PBO.

S-ZDT2 PBO implementations always perform better than the others except for: NSGAII-DE3 with 500 iterations and SPEA2-DE2 with 1000 iterations in which PBO implementations are worse; and NSGAII-DE2 with 500 iterations and SPEA2-DE3 with 1000 iterations in which they are equivalent to the ones based on PBO.

S-ZDT4, S-DTLZ2, R-DTLZ2, S-DTLZ3 In these problems, NSGAII-PBO and SPEA2-PBO present a clearly better behaviour than the others since the triplets are '$(+\,+\,+)$' except for NSGAII-O with 100 iterations in R-DTLZ2 in which a triplet '$(+\;==)$' appears.

R-ZDT4 In this problem, only the implementations based on DE1 and DE2 are the ones that try to confront the PBO implementations, but reaching only equivalent behaviours (it can be observed some triplets '$(=,=,=)$' or '$(=,=,+)$' when 500 and 1000 iterations are considered).

S-ZDT6 For this problem, implementation NSGAII-DE2 with 500 and 1000 iterations is the

Table 3: NSGA-II (top) and SPEA2 (bottom) results of the hypothesis tests performed for the test problems separately.

NSGA-II		S-ZDT1. PBO versus				S-ZDT2. PBO versus				S-ZDT4. PBO versus				R-ZDT4. PBO versus			
No.iter.	Mea.	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3
100	HD	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
500	HD	+	+	+	+	+	+	+	+	+	+	+	+	-	=	=	+
	GD	+	+	+	+	+	+	-	-	+	+	+	+	+	=	=	+
	C	+	+	+	+	+	+	=	-	+	+	+	+	+	=	=	+
1000	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	=	+
	GD	+	+	+	+	+	+	+	+	+	+	+	+	+	=	=	+
	C	+	+	+	+	+	+	=	-	+	+	+	+	+	=	=	+

NSGA-II		S-ZDT6. PBO versus				S-DTLZ2. PBO versus				R-DTLZ2. PBO versus				S-DTLZ3. PBO versus			
No.iter.	Mea.	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3
100	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	+	+	+	+	+	+	=	+	+	+	+	+	+	+
	C	+	+	+	+	+	+	+	+	=	+	+	+	+	+	+	+
500	HD	+	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
1000	HD	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	+	-	=	+	+	+	+	+	+	+	+	+	+	+	+

SPEA2		S-ZDT1. PBO versus				S-ZDT2. PBO versus				S-ZDT4. PBO versus				R-ZDT4. PBO versus			
No.iter.	Mea.	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3
100	HD	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
500	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	=	=	+
	GD	+	+	+	+	+	+	+	+	+	+	+	+	+	=	=	+
	C	+	+	+	+	+	+	=	+	+	+	+	+	+	=	=	+
1000	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	=	=	+
	GD	+	+	+	+	+	+	-	=	+	+	+	+	+	=	=	+
	C	+	+	+	+	+	+	-	=	+	+	+	+	+	=	+	+

SPEA2		S-ZDT6. PBO versus				S-DTLZ2. PBO versus				R-DTLZ2. PBO versus				S-DTLZ3. PBO versus			
No.iter.	Mea.	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3	O	DE1	DE2	DE3
100	HD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
500	HD	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
1000	HD	+	+	=	+	+	+	+	+	+	+	+	+	+	+	+	+
	GD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	C	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+

only one that beats NSGA-II-PBO, with triplets '$(=,-,-)$' and '$(-,-,-)$', for 500 and 1000 iterations, respectively. Besides this, SPEA2-DE1 with 500 iterations shows equivalent behaviour to SPEA2-PBO (triplet '$(=+-)$'). In the remaining implementations and number of iterations NSGAII-PBO and SPEA2-PBO are the best.

As a summary of the previous comments we can establish the following: The original implementations of NSGA-II and SPEA2 are always beaten by the PBO implementation. For the other ones, those based on DE1 only get equivalent results to the PBO in problem S-ZDT1 with 100 iterations. Our implementations beats those based on DE2 in the majority of the cases. Nevertheless, NSGAII-DE2 and SPEA2-DE2 implementations could be considered as equivalent to PBO in problem R-ZDT4, NSGAII-DE2 is better than NSGAII-PBO in problem S-ZDT6 (500 and 1000 iterations) and SPEA2-DE2 beats SPEA2-PBO in S-ZDT2 with 100 iterations. Finally, only in problem S-ZDT2 with 500 iterations, the implementation NSGAII-DE3 beats NSGAII-PBO; in all the other cases NSGAII-PBO and SPEA2-PBO behave better than the ones based on DE3. In the light of these results we can say that implementations based on PBO perform better than the others for the considered set of test problems.

§6. Conclusions

In this work we have presented the implementation of two well-known MOEA's, NSGA-II and SPEA2, with a different set of variation operators: The ones used by the original authors, some based on differential evolution operators and the ones proposed by the authors in [2, 3]. After the experiment carried out we can conclude that the implementations of NSGA-II and SPEA2 that use our operators behave, in general, better than the others.

Acknowledgements

This research work has been supported by the Spanish Ministry of Education and Science under grant MTM2007-66893.

References

[1] ABBAS HA, SARKER R, NEWTON C. PDE: A Pareto-frontier differential evolution approach for multi-objective optimization problems. *Proc. Congress on Evolut. Comput. CEC'2001* (Vol. 2) (2001). Piscataway, New Jersey, IEEE Service Center, 971–978

[2] ALBERTO I AND MATEO PM. Using Pareto Optimality for Defining the Mutation Operator Step Size. Technical Report. *Monografías del Seminario Matemático García Galdeano*, University of Zaragoza, 8, 2008, 1–19.

[3] ALBERTO I AND MATEO PM (2009) A crossover operator that uses Pareto Optimality in its definition. TOP. doi:10.1007/s11750-009-0082-7

[4] CEC (2007) Special Session and Competition on Performance Assessment of Multi-Objective Optimization Algorithms. Available in web page http://www3.ntu.edu.sg/home/epnsugan/. Accessed August 2008

[5] COELLO CA, LAMONT GB AND VAN VELDHUIZEN DA. *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd edn. Springer, Berlin, 2007.

[6] DEB, K. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley and Sons, Chichester, 2001.

[7] DEB K, PRATAP A, AGARWAL S AND MEYARIVAN T A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE *T. Evolut. Comput.* 6(2)(2002a), 182–197.

[8] DEB K, THIELE L, LAUMANNS M AND ZITZLER E. Scalable multi-objective optimization test problems. *Proc. Congress on Evolut. Comput*. CEC'2002 (Vol. 1) (2002b). IEEE Press, Piscataway, New Jersey, pp 825–830

[9] FONSECA CM, PAQUETE L, AND LOPEZ-IBANEZ M (2006) Computation of the Hypervolume Indicator, codes available in web page http://sbe.napier.ac.uk/ manuel/hypervolume. Accessed August 2008

[10] GOLDBERG DE Genetic Algorithms in Search, Optimization and Machine Learning. *Addison Wesley*, Reading, Massachusetts, 1989.

[11] HUANG VL, QIN AK, DEB K, ZITZLER E, SUGANTHAN PN, LIANG JJ, PREUSS M, HUBAND S. Problem definitions for performance assessment on multiobjective optimization algorithms. Tech. Rep. CEC-TR2007, Nanyang Technological University, Singapore, (2007).

[12] IORIO AW AND LI X. Solving Rotated Multi-objective Optimization Problems Using Differential Evolution. *Proc. Joint Austral. Conf. Artif. Intell.*, G. Webb and X. Yu, Eds., Cairns, Australia, 861–872 (2004).

[13] MADAVAN, NK Multiobjective Optimization Using a Pareto Differential Evolution Approach. *Proc. Congress on Evolut. Comput. CEC'2002* (Vol. 2) (2002), 1145–1150. IEEE Service Center, Piscataway, New Jersey.

[14] SCHAFFER JD Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. PhD Thesis, Vanderbilt University, Nashville, Tennessee, (1984).

[15] STORN R AND PRICE K. Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. *Tech. Rep. TR-95-12, International Computer Sicience*, Berkeley, California,(1995) pp 1–12 .

[16] XUE F, SANDERSON AC AND GRAVES RJ. Pareto-based Multi-Objective Differential Evolution. *Proc. Congress on Evolut. Comput. CEC'2003* (Vol. 2)(2003), 862–869

[17] ZITZLER E, DEB K AND THIELE L. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolut. Comput.* 8(2) (2000), 173–195.

[18] ZITZLER E, LAUMANNS M AND THIELE L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Tech. Rep. 103, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institue of Technology (ETH), Zurich, Switzerland, (2001) pp 1–21.

Monografías Matemáticas García de Galdeano **36**, 109–116 (2010)

STOCHASTIC RECURSIVE TECHNIQUES IN RESPONSE-ADAPTIVE DESIGNS

José A. Moler, Fernando Plo and Henar Urmeneta

Abstract. Urn models can be used to generate sequential, response-adaptive designs with good properties. For instance, in the context of clinical trials they can allocate more patients to the best treatment, which is a desirable property from an ethical point of view.

An adaptive design generated by a randomly reinforced urn has been thoroughly studied by May and Flournoy in [8] where its applications to clinical trials are also discussed.

Stochastic recurrence is a powerful tool to obtain asymptotic results for urn models with random replacement matrices (see [6]). Our aim in this paper is to apply these stochastic recurrence techniques to the model studied in [8] in order to obtain similar asymptotic results in a more general setting. By means of a simulation, we also compare an adaptive design based on this model with other well-known adaptive designs.

Keywords: Adaptive designs, Pólya urns, Optimality

AMS classification: 62K99, 62L05, 62L20

§1. Introduction

The classical Pólya urn model consists in an urn which initially contains W_0 white balls and B_0 black balls. The composition of the urn is changed according to the following rule: a ball is drawn and replaced in the urn and $a > 0$ balls of the same color are added to the urn. Let (W_n, B_n) the composition of the urn after n draws. If $T_n = B_n + W_n$ then $X_{1,n} := W_n/T_n$ and $X_{2,n} := B_n/T_n$ represent the proportion of white balls and black balls in the urn, respectively. We denote by $X_n := (X_{1,n}, X_{2,n})$ the vector of proportions. It is well-known, see for instance [1], that $\{X_{1,n}\}$ converges with probability one to a beta distributed random variable $\beta(W_0/a, B_0/a)$.

As we have seen, in the classical Pólya urn model the following two actions are possible in each step: extracting a white ball (action 1) or extracting a black ball (action 2). Let us consider $\boldsymbol{\delta}_n^t = (\delta_{1,n}, \delta_{2,n})$, where $\delta_{i,n}$ is the indicator variable of action i in the nth draw. Now, if we associate the color white with number 1 and the color black with number 2, the diagonal matrix

$$R = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

indicates the replacement process because each component r_{ij} represents the number of balls of color j that are added in the urn when action i happens. Besides, given the sequence of σ-algebras $\{\mathcal{F}_n\}$ where $\mathcal{F}_n = \sigma(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n, X_1, \ldots, X_n)$, the replacement policy is given by

the conditional probability distribution of $\boldsymbol{\delta}_n$, where $\boldsymbol{\delta}_n^t = (\delta_{1,n},\, \delta_{2,n})$. It is straightforward to see that in the classical Pólya urn model

$$P(\boldsymbol{\delta}_n^t = (1,\, 0)|\mathcal{F}_{n-1}) = X_{1,n-1}, \qquad P(\boldsymbol{\delta}_n^t = (0,\, 1)|\mathcal{F}_{n-1}) = 1 - X_{1,n-1}.$$

In the probabilistic literature several generalizations of the previous model have been studied by considering more general replacement matrices, more general replacement policies, or both. It is necessary to impose some restrictions on these models in order to obtain asymptotic results for the proportion of balls in the urn or the proportion of times that each action has been applied, see, for instance, [3], [1], [6] and the references therein. Even if these models are quite general, new generalizations appear that require new techniques to obtain their asymptotic behavior.

In [2] the classical Pólya urn model was generalized in the following way. Consider two sequences of independent and identically distributed random variables, $\{M_n\}$ and $\{N_n\}$. We assume that, for each n, M_n and N_n are positive, bounded by a constant value β and independent of $\boldsymbol{\delta}_n$. Let $E[M_n] = \mu$ and $E[N_n] = \nu$. We consider the classical Pólya urn model and assume that in each step the replacement matrix is

$$R_n = \begin{pmatrix} M_n & 0 \\ 0 & N_n \end{pmatrix}. \tag{1}$$

The proportion of white balls, $\{X_{1,n}\}$, converges to a beta distribution if and only if both variables are degenerated in the same constant value a.

When $\{M_n\}$ and $\{N_n\}$ are non-negative, that is, when they can take the value zero, the asymptotic behavior of this model has been considered again in several papers, see, for instance, [8] and [10]. It is well-known that the proportion of white balls, $\{X_{1,n}\}$, converges to a random variable X_∞, but the distribution of this limit and the speed of convergence are still open problems. In what follows, this model will be called randomized classical Pólya urn model (RCPU).

The paper is organized as follows. In section 2 we apply the techniques used in [6] for generalized Pólya urn models (GPU) to the RCPU model in order to give a new insight to those open problems. In section 3 the RCPU model is used to obtain a response-adaptive design for a clinical trial, comparing its performance, using simulation techniques, with the performances of other well-known response-adaptive designs.

§2. Stochastic recurrence for the randomized classical Pólya urn

It is easy to see that the evolution of the process $\{X_n\}$ in a RCPU model follows a recurrence scheme:

$$\begin{aligned} X_{n+1} &= \frac{T_n X_n + R_{n+1}\boldsymbol{\delta}_{n+1}}{T_{n+1}} \\ &= X_n + \frac{R_{n+1}\boldsymbol{\delta}_{n+1} - (1,1)R_{n+1}\boldsymbol{\delta}_{n+1}X_n}{T_{n+1}}. \end{aligned} \tag{2}$$

Then, the following result is obtained.

Theorem 1. *Consider the randomized classical Pólya urn model (RCPU). The process* $\{X_n\}$ *evolves as the Robbins-Monro recurrence equation*

$$X_{n+1} = X_n + \frac{1}{T_{n+1}}(F(X_n) + \varepsilon_{n+1}).$$

where

- $F(X_n) := RX_n - [(1,\ 1)RX_n]X_n$ *is the Robbins-Monro function, with*

$$R := E[R_n] = \begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix} \tag{3}$$

- $\{\varepsilon_n\}$ *is a sequence of martingale differences.*

Proof. Let $a_{n+1} := R_{n+1}\boldsymbol{\delta}_{n+1} - (1,\ 1)R_{n+1}\boldsymbol{\delta}_{n+1}X_n$, and define

$$\varepsilon_{n+1} := a_{n+1} - E[a_{n+1}|\mathcal{F}_n],$$

for each n. Now, the result follows straightforwardly from (2). □

In order to obtain limit results for a Robbins-Monro process it is crucial to establish the growing order of the step size $\{1/T_n\}$. This is obtained in the next result.

Theorem 2. *Consider the randomized classical Pólya urn model (RCPU). If* $max(\mu,\ \nu) > 0$, *then there exist constants* $K_1 > 0$ *and* $K_2 > 0$ *such that, for any* n

$$K_1 \leq \frac{T_n}{n} \leq K_2.$$

Proof. Let $A_n := (1,\ 1)R_n\boldsymbol{\delta}_n$.
From the hypotheses of the model we have that, for each n, $A_n \in [0,\ \beta]$.
Observe that $T_n = T_0 + \sum_{i=1}^{n} A_i$. Then, we have

$$\begin{aligned} D_n := T_0 + \sum_{i=0}^{n} E[A_i|\,\mathcal{F}_{i-1}] &= T_0 + \sum_{i=0}^{n}(\nu + X_{1i}(\mu - \nu)) \\ &= T_0 + n\nu + (\mu - \nu)\sum_{i=0}^{n} X_{1i} \end{aligned} \tag{4}$$

Without loss of generality we consider that $\mu \geq \nu \geq 0$. The three following cases are possible:

- $\mu > \nu = 0$. Note that $(W_0 + x)/(T_0 + x) \geq W_0/T_0$, for any non-negative real value x. Then

$$T_0 + n\mu\frac{W_0}{T_0} \leq D_n \leq T_0 + n\frac{\beta}{T_0}.$$

- $\mu > \nu > 0$. Then
$$T_0 + n\nu \leq D_n \leq T_0 + n\mu.$$
- $\mu = \nu > 0$. we observe that, for each n, $D_n = T_0 + n\nu$.

So that, there exist constants K_1' and K_2' such that for all n, $K_1' \leq D_n/n \leq K_2'$. Now from Corollary 2.3 in [4] we have that $T_n/D_n \to 1, \quad a.s$, and the result follows. □

Theorem 3. *Consider the randomized classical Pólya urn model (RCPU). If $\mu > \nu$, then the sequence $\{X_n\}$ converges a.s. to $(1,\, 0)$. If $\mu < \nu$, then the sequence $\{X_n\}$ converges a.s. to $(0,\, 1)$.*

Proof. Observe that if a is a positive real constant, the ordinary differential equation

$$\dot{x} = -ax^2 + ax$$

has two stable points, and if $x_0 = 0$, the solution is 0; but if $x_0 > 0$ it converges to 1. From Theorem 5.2.1 in [7] and taking $a = \mu - \nu$ the result follows. □

We denote $\Delta_1 = \{x \in \mathbb{R}^2 \,:\, x_1 + x_2 = 1,\, x_1 \geq 0\ x_2 \geq 0\}$.

Theorem 4. *Consider the randomized classical Pólya urn model (RCPU). If $\mu = \nu$, then F is null for any point in Δ_1. If $\{M_n\}$ and $\{N_n\}$ are identically distributed, then $\{X_{1,n}\}$ is a martingale and converges to a random variable with support in $[0,\, 1]$.*

Proof. From Theorem 2.1, and using that $X_n \in \Delta_1$, we have that

$$\begin{aligned} F(X_n) &= (\mu X_{1,n},\, \nu X_{2,n})^t - [\mu X_{1,n} + \nu X_{2,n}](X_{1,n},\, X_{2,n})^t \\ &= (\mu - \nu) X_{1,n} X_{2,n} (1,\, -1)^t, \end{aligned}$$

so that when $\mu = \nu$ the function F is identically null in Δ_1. Moreover, with some algebra we can obtain that

$$\begin{aligned} X_{1,n} &= X_{1,n-1} + \frac{1}{T_n}\varepsilon_n \\ &= X_0 + \sum_{i=1}^{n} \frac{\varepsilon_i}{T_i} \end{aligned}$$

where $\varepsilon_n := (M_n\delta_{1,n} - \mu X_{1,n-1})X_{2,n-1} - (N_n\delta_{2,n} - \nu X_{2,n-1})X_{1,n-1}$.
When $\mu = \nu$ we have that

$$\begin{aligned} X_{1,n} - X_{1,n-1} &= \frac{M_n\delta_{1,n}X_{2,n-1} - N_n\delta_{2,n}X_{1,n-1}}{T_{n-1} + M_n\delta_{1,n} + N_n\delta_{2,n}} \\ &= \frac{M_n}{T_{n-1} + M_n}\delta_{1,n}X_{2,n-1} - \frac{N_n}{T_{n-1} + N_n}\delta_{2,n}X_{1,n-1}. \end{aligned}$$

Observe that $\boldsymbol{\delta}_n$ is conditionally independent of M_n and N_n, so that

$$E[X_{1,n} - X_{1,n-1}|\mathcal{F}_{n-1}] = E[\frac{M_n}{T_{n-1} + M_n} - \frac{N_n}{T_{n-1} + N_n}]X_{1,n-1}X_{2,n-1},$$

and it follows that $\{X_{1,n}\}$ is a martingale when M_n and N_n are identically distributed. □

Remark 1. When M_n and N_n are not identically distributed the sequence $\{X_{1,n}\}$ may not be a martingale.

§3. Applications in response adaptive designs

We consider a clinical trial where two treatments are compared. The randomized classical Pólya urn model (RCPU) can be used to design the trial. We assume that the patients' responses to each treatment are identically distributed and independent random variables.
Each treatment is associated with a color, say, black or white. We assume that the patients' responses are dichotomous (success or failures). When a patient arrives, a ball is drawn from the urn and this ball is replaced in the urn together with a number of balls equal to the patient's response to the treatment: if the treatment is successful, one ball of the color associated to this treatment is added in the urn, otherwise the urn remains unchanged.
As we have seen in the previous section, when the mean responses to both treatments are not equal, all the patients will eventually be allocated in the best treatment. This is a good property of this design from an ethical point of view. Besides, the asymptotic normality of the difference of sample means, when the randomized classical Pólya urn design (RCPU) is applied in a clinical trial, is proved in [8].
In this section, we present the results of a simulation study carried out to compare the performance of several adaptive-designs, from the ethical and inferential point of view, when the sample size n is small. Our goal is to see the advantages and disadvantages of the RCPU design with respect to other designs.
We consider the hypothesis test:

$$H_0 : \mu = \nu \qquad\qquad H_1 : \mu \neq \nu$$

where μ and ν are the success probabilities for treatment white and black, respectively. We consider the following designs: complete randomization (CR), Play-the-winner rule (PTW) and Drop-the-loser rule (DL). For a complete description of these designs see, for instance, [5] where, in Chapter 8, a comparative study among them can be found.
In this paper we seize table 8.3 in [5] as a benchmark and it is reproduced in Table 1, both Play-the-winner rule and Drop-the-loser rule are initialized with 5 balls of each color in the urn. As stated in [5], the DL rule overcomes the PTW rule because it has better power and the number of failures is equal or smaller in some cases and, as they remark, any value of the decrement in the number of failures is desirable because it means the recovery of a patient.
In Table 2, we show the results of a simulation study for the randomized classical Pólya urn design (RCPU) for the same success probabilities and sample sizes as in Table 1. We denote b and w the initial number of black balls and white balls in the urn, respectively. We have considered three scenarios, $b = w = 1$, $b = w = 3$ and $b = w = 5$.
It is well-known that the complete randomization rule (CR) has the best properties from the inferential point of view because the variability of the test statistic reaches its minimum value. However, from the ethical point of view, this design does not use the information accrued in the trial to allocate the following patient, which is a drawback from an ethical point of view. In fact, there is no drift towards the treatment with the best performance as it happens in response adaptive designs, as PTW, DL and RCPU.

Table 1: **Table 8.3 in [5]** Simulated power and expected treatment failures(s.d) for complete randomization, play the winner rule and drop the loser rule. 10000 replications, $\alpha = 0.05$ two-sided. The sample size was selected that yielded simulated power of approximately 0.9 under complete randomization.

			CR		PTW		DL	
μ	ν	n	power	failures	power	failures	power	failures
0.9	0.3	24	90	10(2.4)	87	7(2.4)	90	7 (1.8)
0.9	0.5	50	90	15(3.2)	87	12(3.2)	89	12 (2.6)
0.9	0.7	162	90	32(5.1)	88	28(5.4)	89	27 (4.6)
0.9	0.8	532	90	80(8)	89	75(9)	89	73 (8)
0.7	0.3	62	90	31(4.0)	88	28(4.3)	89	27 (4.1)
0.7	0.5	248	90	99(7.8)	89	94(8.2)	89	93 (8.0)
0.5	0.4	1036	90	570(16)	89	565(16)	89	565 (16)
0.3	0.1	158	90	126(5.1)	89	125(5.4)	90	124(5.3)
0.2	0.1	532	90	452(8)	89	451(8)	90	451(8)

Table 2: Simulated power and expected treatment failures(s.d) for the randomized classical Pólya urn designs with different initial number of black and white balls, denoted b and w respectively, 10000 replications, $\alpha = 0.05$ two-sided. The sample size was selected that yielded simulated power of approximately 0.9 under complete randomization.

			RCPU (b=w=1)		RCPU (b=w=3)		RCPU (b=w=5)	
μ	ν	n	power	failures	power	failures	power	failures
0.9	0.3	24	79	6(3.1)	85	7(2.8)	87	8(2.6)
0.9	0.5	50	77	11(5)	84	12(3.9)	85	13(3.6)
0.9	0.7	162	70	28(9.4)	82	29(7)	86	29(6.5)
0.9	0.8	532	71	76(17)	83	75(12.2)	87	76(10.9)
0.7	0.3	62	79	25(6)	86	26(4.8)	86	27(4.5)
0.7	0.5	248	72	90(13.8)	83	91(10.7)	85	92(9.8)
0.5	0.4	1036	73	552(31.5)	84	553(23.4)	87	555(21.7)
0.3	0.1	158	83	116(7.1)	88	119(6.5)	89	120(6.4)
0.2	0.1	532	79	436(12.5)	87	439(10.8)	88	442(10.1)

It is clear that ethicality and inferential properties are competitive goals because, for the same sample size, the power of the t-test is greater in the complete randomization design than in the three response-adaptive designs. There is no dramatic loss of power in any case, except when the RCPU is started with $b = w = 1$. On the other hand, the expected number of failures in all the response adaptive designs (PTW, DL and RCPU) is smaller than in the CR designs for the same sample size, and specially when the RCPU is started with $b = w = 1$. The CR design and the RCPU design with $b = w = 1$ are antagonists in the sense that the former is the best one from the inferential point of view but the worst for ethical reasons, and the latter is just the opposite.
The properties of adaptive designs depend heavily on the initial composition of the urn. This dependence is shown in Table 2 for the RCPU design. Both power and number of failures increase when we start the urn with more balls of both colors.
Comparing the RCPU ($b = w = 5$) in Table 2 with PTW and DL in Table 1 (also with $b = w = 5$ as initial values) we can see that the simulated power in the RCPU is slightly smaller than in PTW and DL. Besides, in the first four rows, when the success probability of the first treatment is $\mu = 0.9$, both PTW and DL have a smaller number of failures than RCPU. When $\mu = 0.7$, the number of failures of RCPU is smaller than PTW and it is similar to the number of failures of DL. In the rest of rows, where $\mu \leq 0.5$, RCPU has a smaller number of failures than both PTW and DL.

Acknowledgements

This research was supported by projects MTM2007-63769 and MTM2010-15972 of MICINN. The authors are members of the research group Modelos Estocásticos (DGA).

References

[1] ATHREYA, K. B. AND NEY, P. E. *Branching Processes*. Springer-Verlag, Berlin. 1972.

[2] ATHREYA, K. R. On a characteristic property of Pólya's Urn. *Studia Scientiarum Mathematicarum Hungarica*. **4,** 31–35, 1969.

[3] GOUET, R. Strong Convergence of Proportions in a multicolor Pólya Urn. *Journal of Applied probability*. **34,** 426–435, 1997.

[4] HALL, P. AND HEYDE, C.C. *Martingale limit theory and its applications*. Academic Press, New York. 1980.

[5] HU, F. AND ROSENBERGER, W. F. *The theory of response-adaptive randomization in clinical trials*. Wiley, Hoboken, NJ. 2006.

[6] HIGUERAS, I., MOLER, J. A., PLO, F. AND SAN MIGUEL, M. Central limit theorems for generalized Pólya urn models. *Journal of Applied probability*. **43**, 938–951, 2006.

[7] KUSHNER, H. J. AND YIN, G. G. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York. 1997.

[8] MAY, C. AND FLOURNOY, N. Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Annals of Statistics*. **37**(2), 1058–1078, 2007.

[9] MOLER, J.A., PLO, F. AND URMENETA, H. Optimality for combinations of adaptive and response-adaptive designs. *Monografías del Seminario Matemático García de Galdeano*. **34**, 145–153, 2008.

[10] ZHANG, L., HU, F., CHEUNG, S. H. AND CHAN, W. S. Multi-color randomly reinforced urn for adaptive designs. *arXiv:0812.3699v1Stat.Med* 19th of December of 2008.

José A. Moler and Henar Urmeneta
Departamento de Estadística e Investigación Operativa
Universidad pública de Navarra
Campus Arrosadia s/n, 31006-Pamplona, Spain.
jmoler@unavarra.es and
henar@unavarra.es

Fernando Plo
Departamento de Métodos Estadísticos
Facultad de Ciencias and BIFI
Universidad de Zaragoza
Pedro Cerbuna, 12 50009-Zaragoza, Spain
fplo@unizar.es

Monografías Matemáticas García de Galdeano **36**, 117–126 (2010)

ON THE CHOICE OF SCHOOLS LOCATED OUTSIDE THE WALKABLE NEIGHBOURHOOD OF THE HOUSEHOLD

Augustine A. Osagiede and Virtue U. Ekhosuehi

Abstract. In this study, we consider a school choice problem and formulate it into a mathematical model, allowing it to be simplified and solved. The results obtained are useful for the household in making an objective choice of school for the child to be enrolled among several secondary schools located outside his walkable neighbourhood which best satisfies his budget constraint with emphasis on the most preferred travelling mode of a given type.

Keywords: school choice modelling, Lagrangian function, Stone's utility function, saturation point.

AMS classification: 91B16, 97A80.

§1. Introduction

Earlier studies on school choice modelling [1] and mode of students' travel to school [2] have employed utility functions. Utility functions are well-known in literature see [3] and [4]. Mancebon and Muniz [5] compared the efficiency of a set of Spanish public and private high schools using data envelopment analysis (DEA) and also identified the school inputs.
In this study, we consider the school choice problem of a household H living in an XY community in the Niger Delta region in Nigeria who is willing to enrol his child in basic 7 (formerly referred to as J.S.S. 1) in a standard private-independent secondary school far-off from his residence. Three of such private-independent schools, denoted as $Sch1$, $Sch2$ and $Sch3$, are being considered by the household (see figure 1). In each of these schools the number of teaching staff is exactly equal to the number of subjects taught at that level. The household considered here wishes to enrol his child in a private-independent secondary school based on the available human and physical resources in the school. These include, in descending order of importance as perceived by the household, the following: qualified teaching staff, laboratory, health centre, hostel, computers and sports. The household budgets the sum of one hundred and fifteen thousand naira (₦ 115,000) for a session for the child's education. A survey of the three standard private-independent secondary schools reveals the following attributes and statistics for basic 7 for the academic year as presented in Table 1.
A river separates the household from the nearest motor park (denoted as node 1 in figure 1) where the child can board taxi, bus or motor bike. Speed boats are available for transporting people and their goods across the river at a cost of ₦ 150. The allowable speed limit of the

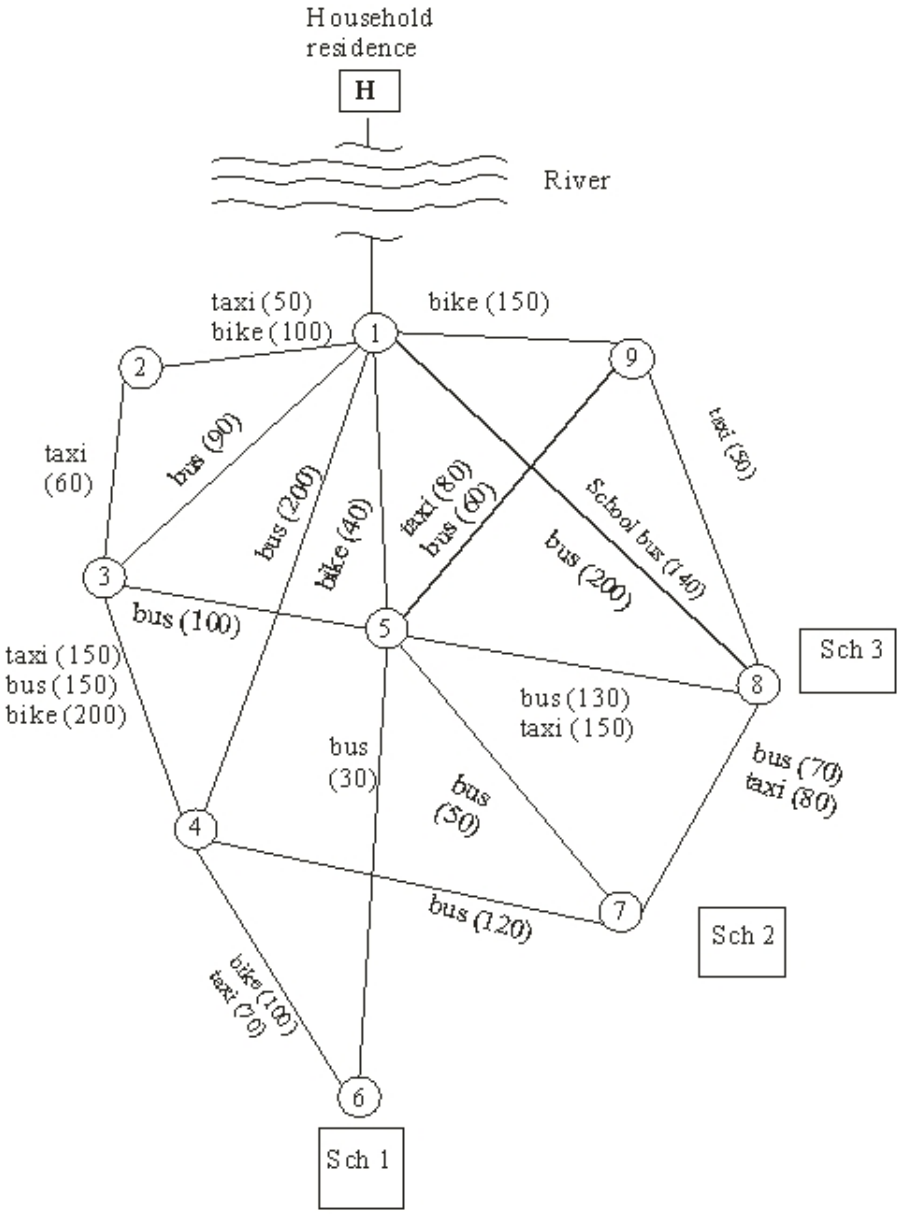


Figure 1: Schematic representation of the transport network and school locations.

School attributes	Fees particulars	School fees (in Naira)			Available input variables per attribute			Maximum student size and distribution for each attribute		
		f_{i1}	f_{i2}	f_{i3}	Sch 1	Sch 2	Sch 3	Sch 1	Sch 2	Sch 3
1. Teaching staff	Tuition and examination fees	40,000	35,000	50,000	12	14	16	40	40	35
2. Hostel	Accommodation fees and living cost	35,000	40,000	45,000	2	2	2	80	60	60
3. Health centre	Medical fees	5,000	10,000	6,500	1	1	1	6	8	6
4. Laboratory	Laboratory fees	10,000	20,000	8,000	1	1	1	20	16	20
5. Computer utilization	Utility charge	4,000	5,000	3,500	6	8	4	20	16	16
6. Sports	Common services charge	2,000	2,500	1,000	3	4	2	22	22	22
		96, 000	112, 500	114, 000						

Table 1: School attributes and vital statistics

boat is 80 km/h. There are eight other nodes linked to node 1. Some of the links are in deplorable state; for this reason, only motor bikes ply such routes.
Each of the three schools is located at node 6, node 7 and node 8. The schematic representation of the network (without the winding routes), mode of transportation and its associated costs (in Naira) enclosed in parenthesis are given in figure 1. For the safety of commuters and pedestrians, buses, cars and bike often do not exceed 50 km/h, 60 km/h and 40 km/h, respectively. The household prefers that the child takes a taxicab (or taxi in a shorter form) to school regardless of the opportunity cost of transportation. The household, therefore, is in a quagmire of not only choosing a secondary school from the three standard private-independent secondary schools but also to determine the number of trips to be made using taxi for the child's trip to school.

§2. Modelling the school choice problem

Since the household seeks a standard private-independent school for the child, we first of all define what we mean here by standard schools as schools where the inputs per capita exceed the minimum specification for accreditation. Although we could not obtain exact information on the accreditation guide by the Ministry of Education in the state, we assume here that the ministry accredits schools if such schools have adequate qualified teaching staff, spacious classroom for at most 40 students per class, standard laboratory equipped for at least practical in the natural and applied science subjects, health centre and computer laboratory. We therefore modify the Stone's utility function as given in [3] by adding the number of trips by travelling modes and its corresponding parameters. By this modification, we develop a model for a household who is willing to enrol his child in a school located outside his walkable neighbourhood which best satisfies his budget constraint with emphasis on the most preferred travelling mode.

2.1. Notations and definitions

In modelling the household school choice problem, we use the following notations. a_i denotes the rank of input variable i according to its preference by the household. n is the maximum number of school input variables under consideration. α_i is the relative weight assigned to input variable i according to its preference by the household. γ_H is amount to be spent on the childŠs education for a session. f_{ij} is the fee charged (per student) for a session for input variable i by school j. $j = 1, \ldots, M$ are the schools under consideration. x_{ij} is the existing input i per capita in school j. x^*_{ij} is the household demand per capita for input variable i in school j. $\beta_i(\geq 0)$is the minimum standard requirement per capita for variable i as specified by the Ministry of Education or any other recognised body for accreditation of schools in the state. $m_q(r, s)$ is the trip from node r to node s by mode m of type q, $\{q = 1, 2, 3, 4, 5\}$, where m stands for commercial engine-powered transport services. m_1 stands for speed boat, m_2 for bike, m_3 for taxi, m_4 for bus, and m_5 for school bus of $Sch3$. $c^q_{r_H s_j}$ is the average cost of transportation per student commuter from node r near H to node s (before reaching school j) using a mode of type q. $m_q \in Q = \bigcup_{q=1}^{p} \{m_q\}$ which is the set

of all transport modes. $A(r,s)$ is the set of transport modes plying route (r,s). $T^{*}_{r_H s_j}$ is the alternative number of trips from the node r near H to the node s, before reaching school j using modes of type q which give the minimum transport cost. ${}^{k}T^{q}_{r_H s_j}$ is the feasible number of trips with minimum cost using mode of type q with emphasis on using the most preferred travelling mode of type k. $T^{k^*}_{r_H s_j}$ is the Lagrange determined number of trips using the most preferred mode k at the household saturation point. $T^{k^{**}}_{r_H s_j}$ is the household expected number of trips using the most preferred mode k. $T^{q'}_{r_H s_j}$ is the feasible number of trips with minimum cost using mode of type q', $m_{q'} \in Q/\{m_k\}$. $\wp_q$ is the proportion of recommended safety speed limit for mode of type q. The most preferred travelling mode is m_3.

2.2. The budget constraint

Since the household has allocated a fixed amount for the child's education, we assume that the household is willing to choose the best private-independent secondary school located outside the walkable neighbourhood of his residence for the child provided the total expenditure on the child's education for the session does not exceed his budget. We also assume that each mode type returns via the route it took on the outgoing trip (tour). Thus, if the amount to be spent on the childŠs education for a session is γ_H, then the household budget constraint is:

$$\left(\begin{array}{c}\text{total cost of school}\\ \text{input variables}\end{array}\right) + \left(\begin{array}{c}\text{total cost of}\\ \text{transportation}\end{array}\right) \leq \gamma_H. \tag{1}$$

Equation 1 can be rewritten as

$$\sum_{i=1}^{n} f_{ij}x_{ij} + 2\sum_{q=1}^{p} c^{q}_{r_H s_j} T^{q}_{r_H s_j} \leq \gamma_H, \tag{2}$$

$x_{ij} \geq 0, T^{q}_{r_H s_j} \geq 0, j = 1, \ldots, M.$
The transport mode m of type q may contribute more than once to the total number of trips. Of note is that the existing input i per capita in school j can be computed as follows: for $i = 1$ i.e. teaching staff,

$$x_{1j} = \frac{\text{number of teaching staff in } schj}{(\text{number of subjects taught in } schj) \times (\text{maximum allowableclass size per subject in } schj)}$$

and $x_{ij} = \dfrac{\text{available inputs for attribute } i \text{ in } schj}{\text{maximum student size for attribute } i \text{ in } schj}$, for $i = 2, \ldots, 6.$

2.3. The household utility function

To construct the household utility function, we take $(x_{ij} - \beta_i) > 0 \; \forall i$, as the schools under consideration are standard schools. Expressing the household utility function for school inputs as analogous to Stone's utility function [3], we have

$$U^{j}_{1} = \sum_{i=1}^{n} \alpha_i \log_e(x_{ij} - \beta_i), (j = 1, \ldots, M). \tag{3}$$

To ensure that $\sum_{=1}^{n} \alpha_i = 1$ (which is a constraint for the marginal budget shares in the Stone's utility function), we define the relative weight of preferences α_i as

$$\alpha_i = \frac{a_i}{n(n+1)}. \tag{4}$$

The ranking of the household's preferences to obtain a_i follows from the indifference-curves theory [3]. The weights α_i assigned to the rank of preferences a_i in equation (4) are obtained as follows: recall that the sum of ranks in the linear rank statistic is $\sum_{i=1}^{n} a_i = \frac{n(n+1)}{2}$, see [6]; then divide each rank a_i by the sum of ranks to obtain its relative weight α_i. We also assume that the utility from travelling is given by the number of trips weighted by the proportion of speed limits as

$$U_2^j = \sum_{q=1}^{p} \wp^q T^q_{r_H s_j}, (j = 1, \ldots, M). \tag{5}$$

Assuming that the utility from school inputs and that of travelling are independent, then the household utility function, U_H, is $U_H = U_1^j + U_2^j$, which can be rewritten as

$$U_H = \sum_{i=1}^{n} \alpha_i \log_e(x_{ij} - \beta_i) \sum_{q=1}^{p} \wp^q T^q_{r_H s_j}, (j = 1, \ldots, M). \tag{6}$$

The implication of the assumption leading to the utility function in equation (6) is that there is no possibility of substitution between the school inputs and the number of trips.

2.4. Mathematical representation of the problem

Let F_j denote the total fees charged by school j. The schools $(j = 1, \ldots, M)$ considered by the household are those satisfying the relation

$$F_j < \gamma_H \tag{7}$$

If $M = 1$ then the household has no alternative school. The school choice problem arises when $M > 1$. Since $M > 1$ for the problem under consideration, then the household has to identify which of the schools best satisfies his demands per capita. To determine the household choice of school therefore, we need to provide a match between the household H demand per capita and each available school input variable i in school j as well as that of the number of trips. In line with the foregoing, we develop a model for the household H school choice problem as:

$$\left.\begin{array}{ll} \text{Maximize} & \\ & U_H = \sum_{i=1}^{n} \alpha_i \ln(x_{ij} - \beta_i) + \sum_{q=1}^{p} \wp^q T^q_{r_H s_j} \\ \text{subject to} & \\ & \sum_{i=1}^{n} f_{ij} x_{ij} + 2 \sum_{q=1}^{p} c^q_{r_H s_j} T^q_{r_H s_j} \leq \gamma_H \\ & x_{ij} \geq 0, T^q_{r_H s_j} \geq 0, j = 1, \ldots, M \end{array}\right\} \tag{8}$$

2.5. Deriving solutions from the model

In this sub-section, we present the method for choosing the most desirable school for the household via the theoretic solutions from problem (8) and the transport network. The transport network in figure 1 is a graph of the form $G = (V, E)$ where $r, s \in V$, is the set of nodes, including the household residence, and E is the set of routes. For every route $e = (r, s) \in E \subseteq V \times V$, a nonnegative cost per commuter passing through the route $r \to s$ by a mode of type q, $c[q; r, s]$, is associated with it. There is only one source, H, which is the household residence and several sink nodes, s_j, which are the school locations. Now we develop the routines for finding the number of trips with minimum cost and the number of trips with preference for the most preferred travelling mode of type k as follow.

Step 0: Identify the source, H, and the sinks, s_j, $j = 1, \ldots, M$.

Step 1: Consider each node j^*, directly linked to H. Obtain $c[q^*; H, j^*] = \min\limits_{\forall m_q}\{c[q; H, j^*]\}$ for each $e = (H, j^*) \in E$, and for the most preferred travelling mode of type k,

$$c_k[q^*; H, j^*] = \begin{cases} c[k; H, j^*] & \text{if } \mathrm{m_k} \in \mathrm{A(H, j^*)} \text{ for each } \mathrm{e = (H, j^*)} \in \mathrm{E} \\ \min\limits_{\forall m_{q'} \in Q/\{m_k\}} \{c[q'; H, j^*]\} & \text{if } \mathrm{m_k} \notin \mathrm{A(H, j^*)} \text{ for each } \mathrm{e = (H, j^*)} \in \mathrm{E} \end{cases}$$

(Ties are broken arbitrarily). Set $\zeta = 2$.

General step ζ: Let ψ_{j^*} be a sub-graph of G rooted at node j^*. Let $e^{(\tau)} = (r^{(\tau)}, s^{(\tau)})$, $\tau = 1, \ldots, \omega$, be all possible routes leading to s_j. Compute

$$C[r, s_j] = c[q^*; H, j*] + \min_{\forall m_q} \left\{ \sum_{e \in \Delta} c[q; r, s] \right\}, \quad \text{for each node j}^* \tag{9}$$

where $C[r_H, s_j]$ is the minimum outgoing total cost for the child's trip to school j, and $\Delta = \psi_{j^*} \cap \left(\bigcup\limits_{\tau=1}^{\omega} \{e^{(\tau)}\} \right)$. For the most preferred travelling mode of type k, compute

$$C_k[r_H, s_j] = c_k[q^*; H, j^*] + \min_{\forall m_q} \left\{ \sum_{e \in \Delta} c_k[q; r, s] \right\}, \quad \text{for each node j}^* \tag{10}$$

where

$$c_k[q; r, s] = \begin{cases} c[k; r, s] & \text{for } \mathrm{m_k} \in \mathrm{A(r, s)} \\ c[q'; r, s] & \text{for } \mathrm{m_k} \notin \mathrm{A(r, s)} \end{cases}$$

If $j = M$, stop.

The sequence $\left\{\sigma^{(\tau)}\right\}_{\tau=1}^{\eta}, \eta \leq \omega, \ni s^{(\tau)} = r^{(\tau+1)}, \sigma^{(\tau)} \in \bigcup\limits_{\tau=1}^{\omega} \{e^{(\tau)}\}$, which yields the minimum outgoing total cost $C[r_H, s_j]$, is the path with the minimum cost to school j. The number of trips with minimum cost to school j is therefore: $T^{q^*}_{r_H s_j} = \eta$.

Number of trips using the most preferred travelling mode of type k to school j is:

$${}^{k}T^{k}_{r_H s_j} = {}^{k}T^{q^*}_{r_H s_j} - {}^{k}T^{q'}_{r_H s_j},\ m_{q'} \in Q/\{m_k\}, \tag{11}$$

where ${}^{k}T^{q^*}_{r_H s_j}$ is the number of trips which yields the outgoing total cost $C_k[r_h, s_j]$, and $c^{k^*}_{r_H s_j}$ is the average cost of transportation per student commuter with emphasis on using a mode of type k.

We maximize problem (8) for the case where the constraint takes the sign of equality (i.e. at the household saturation point). The upper bound of the expenditure is then achieved by defining a variable z_j as $z_j = \min(T^{k^*}_{r_H s_j}, {}^kT^k_{r_H s_j})$. Using the Lagrangian method, the Lagrangian function is

$$L = \sum_{i=1}^{n} \alpha_i \log_e(x_{ij} - \beta_i + \sum_{q=1}^{p} \wp^q T^q_{r_H s_j} + \lambda \left[\gamma_H - \sum_{i=1}^{n} f_{ij} x_{ij} - 2 \sum_{q=1}^{p} c^q_{r_H s_j} T^q_{r_H s_j} \right], \quad (12)$$

where λ is the Lagrange multiplier [7]. Differentiating L with respect to x_{ij}, $T^k_{r_H s_j}$,and λ, and solving the resulting equations after setting the derivatives to zero, we obtain

$$x^*_{ij} = \frac{2\alpha_i c^{k^*}_{r_H s_j}}{f_{ij} \wp^k}, (j = 1, \ldots, M), \quad (13)$$

and

$$T^{k^*}_{r_H s_j} = \frac{1}{2c^{k^*}_{r_H s_j}} \left\{ \gamma_H - 2 \left(\frac{c^{k^*}_{r_H s_j}}{\wp^k} + \frac{1}{2} \sum_{i=1}^{n} f_{ij} \beta_i + \sum_{\substack{q=1 \\ q \neq k}}^{p} c^{q'}_{r_H s_j} T^{q'}_{r_H s_j} \right) \right\}, (j = 1, \ldots, M). \quad (14)$$

Equation (13) is the household demand per capita for input variable i in school j, while equation (14) is the Lagrange determined number of trips from node r near H to node s, to be made using the most preferred mode k at the household saturation point for the child's trip to school j. Since x^*_{ij} is expressed as input per capita, it takes fractional values. However, if the child is a day student, then $c^q_{r_H s_j}$ is replaced by $\rho c_{r_H s_j}$, where ρ is the number of school days (excluding holidays) in a session. We match the household H demand per capita for each of the school input variable i and the expected number of trips from node r to node s, made using the most preferred mode k before reaching school j as: $\min_j(\Theta'_j, \Theta_j)$, where $\Theta'_j = (\theta_{1j}, \theta_{1j}, \ldots, \theta_{nj}, \theta_{(k)j})$ is the transposition of Θ_j which is an $(n+1)-$dimensional row vector, θ_{ij} is given as

$$\theta_{ij} = \frac{1}{\sqrt{n+1}} \left| \frac{x_{ij} - x^*_{ij}}{x^*_{ij}} \right| \quad (15)$$

and $\theta_{(k)j}$ is

$$\theta_{(k)j} = \frac{1}{\sqrt{n+1}} \left| \frac{T^k_{r_H s_j} - T^{k^{**}}_{r_H s_j}}{T^{k^{**}}_{r_H s_j}} \right|. \quad (16)$$

(The scaling factor $\sqrt{n+1}$ is employed so as to account for the number of entries in vector Θ_j). $T^{k^{**}}_{r_H s_j}$ is obtained as

$$T^{k^{**}}_{r_H s_j} = [z_j], \quad (17)$$

where $[z_j]$ is the largest integer $\leq z_j$. This is done so as to avoid situations where $T^{k*}_{r_H s_j} >^k T^k_{r_H s_j}$ and the possibility of having fractional values of $T^{k^*}_{r_H s_j}$. $(\Theta'_j \Theta_j)$ measures the extent to which school j satisfies the household demands. If $(\Theta'_j \Theta_j) \to 0$, then school j is a perfect school choice for the household; otherwise the school j is the least desirable school for the household. Hence, the best school choice for the household among all potentially competitive schools is as given by the expression: $\min_j(\Theta'_j, \Theta_j)$. We obtain the household H most desirable school in the next session.

§3. Results and discussion

We implement the algorithm in sub-section 2.5 in MATLAB. To obtain the verisimilitude of decision variables, we first compute the existing input per capita in each of the schools. Thus, we have the following results. For $Sch1$:
$\theta_{11} = 0.0578$, $\theta_{21} = 3.2822$, $\theta_{31} = 0.9054$, $\theta_{41} = 0.0938$, $\theta_{51} = 2.4526$, $\theta_{61} = 3.0454$, $\theta_{(3)1} = 0.3780$.
Thus vector Θ_1 is $\Theta_1 = (0.0578, 3.2822, 0.9054, 0.0938, 2.4526, 3.0454, 0.3780)$ and $\Theta_1'\Theta_1 = 27.0369$. Similarly, we obtain for $Sch2$ and $Sch3$ respectively as: $\Theta_2'\Theta_2 = 45.30$ and $\Theta_3'\Theta_3 = 11.47$. The most desirable school for household H is determined by: $\min_{j=1,2,3}(\Theta_j'\Theta_j) = \Theta_3'\Theta_3 = 11.47$, and the order of desirability is

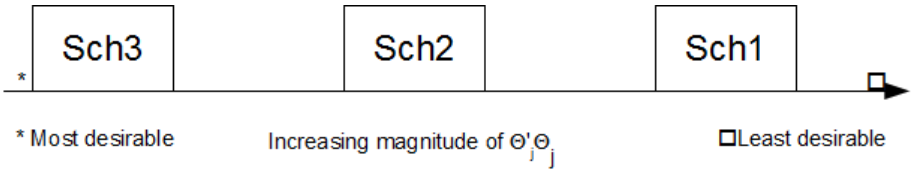


Figure 2: Order of desirability

The result above implies that household H should enrol his child in $Sch3$. The decision to do this is not motivated by the provision of school bus by $Sch3$, as the household prefers that the child takes taxi to school, or because the trip using taxi to $Sch3$ is the cheapest relative to that of the other schools, as the household has sufficient funds; rather it is based on the extent to which each of the three schools is able to meet the household taste and preferences. Enrolling the child in $Sch3$, the household will incur a total cost of ₦ 114,640, and the shortfall from γ_H is the amount to be given to the child as pocket money, which is: pocket money $= \gamma_H -$ ₦ 114,680 $=$ ₦ 320. The child's mode-trip pattern to $Sch3$ is of the form: $m_1(H,1) \rightarrow m_2(1,5) \rightarrow m_3(5,8)$. Observe that although $Sch1$ is the cheapest of the three schools, it is not selected as the most desirable school for the household by the model. By this deduction, we recommend that when a household is faced with the problem of choosing a school among several private-independent schools, the most desirable school should not be determined by the school charging the lowest fees among them, but by the capacity of the school to meet his demands per capita. This study therefore provides a decision-support tool for the household in making a rational choice of school for the child.

§4. Conclusion

This work is a development of a normative prototype for school choice decision which enables the household to choose a school located outside his walkable neighbourhood which best satisfies his budget constraint for the school age child in an objective manner. It also employs existing school input variables and most preferable mode of travelling for school choice rather than the ownership-type as in [1]. Further, it introduces minimum standard of input requirements for schools which directly relate to the child's educational development. Among three private-independent secondary schools considered in this study denoted as $Sch1$, $Sch2$, and $Sch3$, was found to be the most desirable school for the household.

References

[1] C. R. BELFIELD. Modelling School Choice: A Comparison of Public, Private-Independent, Private-Religious and Home-Schooled Students. *Education Policy Analysis Archives*, 12(30) (2004). Retrieved from http://epaa.asu. edu/epaa/v12n30/.

[2] R. EWING, W. SCHROEER AND W. GREENE. School Location and Student Travel: Analysis of Factors Affecting Mode Choice. *Journal of the Transportation Research Board*, 1895(2004), 55–63.

[3] A. KOUTSOYIANNIS. *Modern Microeconomics*. MacMillan Press Ltd., London, 1979.

[4] H. R. VARIAN. *Intermediate Microeconomics: A Modern Approach (6th Ed)*. W. W. Norton and Company, Inc., New York, 2003.

[5] M. J. MANCEBON AND M. A. MUNIZ. Private versus Public High Schools in Spain: Disentangling Managerial and Programme Efficiencies. *Journal of the Operational Research Society, Vol. 59*, No. 7 (2008) 892–901.

[6] B. W. LINDGREN. *Statistical Theory (4th Ed.)*. Chapman and Hall, New York, 1993.

[7] M. I. KAMIEN AND N. L. SCHWARTZ. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management (2nd Ed.)*. Elsevier, Amsterdam, 1992.

Augustine A. Osagiede and Virtue U. Ekhosuehi
Department of Mathematics
University of Benin, Benin City, Nigeria
austaide2006@yahoo.com and myvectorspace2000@yahoo.com

Monografías Matemáticas García de Galdeano **36**, 127–136 (2010)

PRELIMINARY PHI-DIVERGENCE TEST ESTIMATORS IN A CONTINGENCY TABLE WITH SYMMETRY STRUCTURE

Leandro Pardo and Nirian Martín

Abstract. For the model of symmetry in a two way contingency table, shrinkage estimators based on minimum ϕ-divergence estimators and ϕ-divergence test statistics are considered. These estimators are based on the James-Stein type rule and incorporate the idea of preliminary test estimator too. The asymptotic bias and risk are obtained under null and contiguous alternative hypotheses.

Keywords: Minimum Phi-divergence estimator, phi-divergence statistics, Preliminary test estimator, Symmetry model.

AMS classification: 62B10, 62H15.

§1. Introduction

Let X and Y denote two categorical response variables, X and Y having I levels. When we classify subjects on both variables, there are I^2 possible combinations of classifications. The responses (X, Y) of a subject randomly chosen from some population have a probability distribution. Let $p_{ij} = \Pr(X = i, Y = j)$, with $p_{ij} > 0$, $i, j = 1, ..., I$ and we denote by $\boldsymbol{p} = (p_{11}, ..., p_{II})^T$ the joint distribution of X and Y. We display this distribution in a rectangular table having I rows for the categories of X and I columns for the categories of Y. Consider a random sample of size n, $(X_1, Y_1), ..., (X_n, Y_n)$ from (X, Y) and and we denote

$$N_{ij} = \sum_{l=1}^{n} I_{\{i,j\}}(X_l, Y_l) \tag{1}$$

and n_{ij} a particular result of N_{ij}, i.e., n_{ij} represents the observed frequency in the (i, j)th cell for $(i, j) \in I \times I$ with $\sum_{i=1}^{I} \sum_{j=1}^{I} n_{ij} = n$. We shall denote $\widehat{\boldsymbol{p}} = (n_{11}/n, ..., n_{II}/n)^T$. The classical problem of symmetry in a contingency table consists in testing

$$H_0 : p_{ij} = p_{ji}, \quad (i, j) \in I \times I \quad \text{versus} \quad H_1 : p_{ij} \neq p_{ji}, \text{ for at least one } (i, j) \text{ pair.} \tag{2}$$

It was considered for the first time by Bowker (1948). Bowker used the Pearson's statistic

$$X^2 = \sum\nolimits_{\substack{i,j \\ i<j}} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \tag{3}$$

which for large n has a chi-square distribution with $I(I-1)/2$ degrees of freedom. For this problem the likelihood ratio test statistic is given by

$$G^2 = 2\sum\nolimits_{\substack{i,j \\ i<j}} n_{ij} \log \frac{2n_{ij}}{n_{ji} + n_{ij}}. \tag{4}$$

and its asymptotic distribution coincides with the asymptotic distribution of X^2.
We consider the set Θ, defined by

$$\left\{\boldsymbol{\theta}:\boldsymbol{\theta}=(p_{ij};\ 1\leq i\leq I,\ 1\leq j\leq I,\ (i,j)\neq(I,I)),p_{ij}>0,\sum_{i=1}^{I}\sum_{\substack{j=1\\(i,j)\neq(I,I)}}^{I}p_{ij}<1\right\} \tag{5}$$

and we denote $\boldsymbol{p}(\boldsymbol{\theta})=(p_{11},...,p_{II})^T$, $p_{II}=1-\sum_{i=1}^{I}\sum_{\substack{j=1\\(i,j)\neq(I,I)}}^{I}p_{ij}$.
We denote $\boldsymbol{a}=(a_{11},...,a_{1I},a_{22},...,a_{2I},...,a_{I-1I-1},a_{I-1I})^T$ and define

$$B=\{\boldsymbol{a}\in\mathbb{R}^{\frac{I(I+1)}{2}-1}:\sum_{i\leq j}a_{ij}<1,\ 0<a_{ij},\ i,j=1,..,I\}.$$

Hypothesis (2) can be written as

$$H_0:\boldsymbol{\theta}=\boldsymbol{g}(\boldsymbol{\beta}),\quad \boldsymbol{\beta}=(p_{11},...,p_{1I},p_{22},...,p_{2I},...,p_{I-1I-1},p_{I-1I})^T\in B \tag{6}$$

where function $\boldsymbol{g}$ is defined by $\boldsymbol{g}=(g_{ij};\ i,j=1,...,I,\ (i,j)\neq(I,I))$ with

$$g_{ij}(\boldsymbol{\beta})=\left\{\begin{array}{ll}p_{ij} & i\leq j\\ p_{ji} & i>j\end{array}\right.,\quad i,j=1,...,I-1,$$

and $g_{Ij}(\boldsymbol{\beta})=p_{jI}$, $j=1,...,I-1$; $g_{iI}(\boldsymbol{\beta})=p_{iI}$, $i=1,...,I-1$. Note that $\boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))=(g_{ij}(\boldsymbol{\beta});\ i,j=1,...,I)^T$, where

$$g_{II}(\boldsymbol{\beta})=1-\sum_{\substack{i,j=1\\(i,j)\neq(I,I)}}^{I}g_{ij}(\boldsymbol{\beta}).$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ can be defined as

$$\widehat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}\in B}D\left(\widehat{\boldsymbol{p}},\boldsymbol{p}\left(\boldsymbol{g}(\boldsymbol{\beta})\right)\right)\ \text{a.s.}$$

where $D\left(\widehat{\boldsymbol{p}},\boldsymbol{p}\left(\boldsymbol{g}(\boldsymbol{\beta})\right)\right)$ is the Kullback-Leibler divergence measure defined by

$$D\left(\widehat{\boldsymbol{p}},\boldsymbol{p}\left(\boldsymbol{g}(\boldsymbol{\beta})\right)\right)=\sum_{i=1}^{I}\sum_{j=1}^{I}\widehat{p}_{ij}\log\frac{\widehat{p}_{ij}}{g_{ij}(\boldsymbol{\beta})}.$$

We denote by $\widehat{\boldsymbol{\theta}}=\boldsymbol{g}(\widehat{\boldsymbol{\beta}})$ and by $\boldsymbol{p}(\widehat{\boldsymbol{\theta}})=\left(p_{11}(\widehat{\boldsymbol{\theta}}),...,p_{II}(\widehat{\boldsymbol{\theta}})\right)^T$. It is well-known that $\boldsymbol{p}_{ij}(\widehat{\boldsymbol{\theta}})=\frac{\widehat{p}_{ij}+\widehat{p}_{ji}}{2}$, $i=1,...,I$, $j=1,...,I$. Using the ideas developed in Morales et al. (1995), we can consider the minimum ϕ_2-divergence estimator ($M\phi_2E$) replacing the Kullback-Leibler divergence by a ϕ_2-divergence measure in the following way

$$\widehat{\boldsymbol{\beta}}^{\phi_2}=\arg\min_{\boldsymbol{\beta}\in B}D_{\phi_2}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}\left(\boldsymbol{g}(\boldsymbol{\beta})\right)\right);\ \phi_2\in\Phi^*, \tag{7}$$

where

$$D_{\phi_2}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}\left(\boldsymbol{g}(\boldsymbol{\beta})\right)\right)=\sum_{i=1}^{I}\sum_{j=1}^{I}g_{ij}(\boldsymbol{\beta})\phi_2\left(\frac{\widehat{p}_{ij}}{g_{ij}(\boldsymbol{\beta})}\right),$$

Φ^* is the class of all convex functions $\phi_2(x)$, $x > 0$, such that at $x = 1$, $\phi_2(1) = 0$, $\phi_2''(1) > 0$, and at $x = 0$, $0\phi_2(0/0) = 0$ and $0\phi_2(p/0) = p\lim_{u\to\infty}\phi_2(u)/u$. The ϕ_2 -divergence measures were introduced simultaneously by Ali and Silvey (1966) and Csiszàr (1963). For more details about ϕ-divergence measures see Pardo (2006) and references therein. In the following we shall assume that the functions ϕ_2 in the class Φ^* are twice continuously differentiable at $x > 0$.

We denote by $\widehat{\boldsymbol{\theta}}^{\phi_2} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{\phi_2})$ and by

$$\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2}) = \left(p_{11}(\widehat{\boldsymbol{\theta}}^{\phi_2}), ..., p_{II}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)^T \tag{8}$$

the $M\phi_2 E$ of the probability vector that characterizes the symmetry model. Based on $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ it is possible to define a new family of statistics for testing (2) that contains as a particular case the statistics given in (3) and (4). This family of statistics is given by

$$T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2}) \equiv \frac{2n}{\phi_1''(1)} D_{\phi_1}\left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right) = \frac{2n}{\phi_1''(1)} \sum_{i=1}^{I}\sum_{j=1}^{I} p_{ij}(\widehat{\boldsymbol{\theta}}^{\phi_2})\phi_1\left(\frac{\widehat{p}_{ij}}{p_{ij}(\widehat{\boldsymbol{\theta}}^{\phi_2})}\right). \tag{9}$$

We can observe that the family (9) involves two functions ϕ_1 and $\phi_2 \in \Phi^*$. We use the function ϕ_2 to obtain the $M\phi_2 E$ and ϕ_1 to obtain the family of statistics. The asymptotic distribution of $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ is chi-squared with $m = I(I-1)/2$ degrees of freedom (see Chapter 8 in Pardo (2006)). Thus, for a given level of significance $\alpha \in (0,1)$, the critical value of $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ may be approximated by $\chi^2_{m,\alpha}$, the upper $100\alpha\%$ of the chi-square distribution with m degrees of freedom. If we consider in (9), $\phi_2(x) = x\log x - x + 1$ we get the Kullback-Leibler divergence and therefore the corresponding $M\phi_2 E$ is the MLE. If in addition we consider $\phi_1(x) = x\log x - x + 1$ or $\phi_1(x) = (x-1)^2/2$ we obtain the test statistics given in (3) and (4) respectively.

When the hypothesis of symmetry holds, $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ has a smaller risk (with a quadratic loss) than $\widehat{\boldsymbol{p}}$. If the hypothesis of symmetry does not verify, the risk of $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ may go to $+\infty$, as the sample size n increases. For this reason, when the prior knowledge about the hypothesis of symmetry in (2) is rather uncertain, it may be desirable to use a preliminary test estimator. We shall consider in this paper the preliminary phi-divergence test estimator, defined by

$$\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2}) = \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})\mathrm{I}_{\left(0,\chi^2_{m,\alpha}\right)}(T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})) + \widehat{\boldsymbol{p}}\mathrm{I}_{\left[\chi^2_{m,\alpha},\infty\right)}(T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})). \tag{10}$$

Preliminary test estimation (PTE) was introduced by Bancroft (1944). Since then many papers studying the behavior of this procedure of estimation have been published. The book of Saleh (2006) is a good example of the importance of this area of researching.

In Section 2 some asymptotic distributional results are given and in Section 3 we present the asymptotic bias as well as the asymptotic distributional quadratic risk for $\widehat{\boldsymbol{p}}$, $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ and $\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})$.

§2. Asymptotic distributional results

The Fisher information matrix of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a $\left(I^2-1\right) \times \left(I^2-1\right)$ matrix given by

$$\boldsymbol{I}_F^S(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \boldsymbol{H}(\boldsymbol{\theta})^T \left(\boldsymbol{H}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \boldsymbol{H}(\boldsymbol{\theta})^T \right)^{-1} \boldsymbol{H}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{\boldsymbol{\theta}},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = diag(\boldsymbol{\theta}) - \boldsymbol{\theta}\boldsymbol{\theta}^T$ and $\boldsymbol{H}(\boldsymbol{\theta}) = \left(\frac{\partial h_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{ij}} \right)_{\frac{I(I-1)}{2} \times (I^2-1)}$. The functions h_{ij}, are given by

$$h_{ij}(\boldsymbol{\theta}) = p_{ij} - p_{ji}, \qquad i < j, i = 1, ..., I-1, j = 1, ..., I.$$

For more details see Chapter 8 in Pardo (2006). It is not difficult to establish that the Fisher information matrix $\boldsymbol{I}_F^S(\boldsymbol{\theta})$ can be written as $\boldsymbol{I}_F^S(\boldsymbol{\theta}) = \boldsymbol{M}_{\boldsymbol{\beta}}^T \boldsymbol{I}_F(\boldsymbol{\beta})^{-1} \boldsymbol{M}_{\boldsymbol{\beta}}$, where $\boldsymbol{I}_F(\boldsymbol{\beta})$ is the Fisher information matrix corresponding to $\boldsymbol{\beta} \in B$.

We consider a contiguous sequence of alternative hypotheses that approaches the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))$, for some unknown $\boldsymbol{\beta} \in B$, at the rate $O\left(n^{-1/2}\right)$. Consider the multinomial probability vector $p_{n,ij} = \boldsymbol{p}_{ij}(\boldsymbol{g}(\boldsymbol{\beta})) + d_{ij} n^{-1/2},\ i = 1, ..., I,\ j = 1, ..., I$, where $\boldsymbol{d} = (d_{11}, ..., d_{II})^T$ is a fixed $I^2 \times 1$ vector such that $\sum_{i=1}^I \sum_{j=1}^I d_{ij} = 0$, recall that n is the total count parameter of the multinomial distribution and $\boldsymbol{\beta} \in B$. As $n \to \infty$, the sequence of multinomial probabilities $\{\boldsymbol{p}_n\}_{n \in N}$ with $\boldsymbol{p}_n = (p_{n,ij},\ i = 1, ..., I,\ j = 1, ..., I)^T$, converges to a multinomial probability in H_0 at the rate of $O\left(n^{-1/2}\right)$. Let

$$H_{1,n} : \boldsymbol{p}_n = \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})) + \boldsymbol{d} n^{-1/2}, \qquad \boldsymbol{\beta} \in B \tag{11}$$

be a sequence of contiguous alternative hypotheses, here contiguous to the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))$, for some unknown $\boldsymbol{\beta} \in B$. We can observe that $\boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))$ with $\boldsymbol{\beta} \in B$ is given by $\boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})) = (p_{ij}, i, j = 1, ..., I; p_{ij} = p_{ji})^T$. We shall denote

$$\Theta_0 = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\beta}) \text{ for some } \boldsymbol{\beta} \in B\}.$$

It is not difficult to obtain the asymptotic distribution of $\boldsymbol{p}_{\phi_1}^{\text{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})$, i.e., the asymptotic density function of $\boldsymbol{W}_n = \sqrt{n} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2} \left(\boldsymbol{p}_{\phi_1}^{\text{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2}) - \boldsymbol{p}(\boldsymbol{\theta}) \right)$, under $H_{1,n}$. The density function $f_{\mathbf{W}}(\mathbf{x})$ of $\boldsymbol{W}_n$ is given by

$$\phi_{N(\mathbf{0}, \mathbf{K}(\boldsymbol{\theta}))}(\mathbf{x} - \boldsymbol{K}(\boldsymbol{\theta}) \boldsymbol{\delta}) \, G_m\left(\chi_{m,\alpha}^2; \Delta\right) + \int_{\mathrm{E}[\boldsymbol{\delta}]} \phi_{\mathcal{N}(\mathbf{0}, \mathbf{K}(\boldsymbol{\theta}))}(\mathbf{x} - \boldsymbol{\delta} - \mathbf{t}) \, \phi_{\mathcal{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}))}(\mathbf{t}) \, d\mathbf{t},$$

where $\phi_{\mathcal{N}(\mathbf{0}, \boldsymbol{A})}(\mathbf{x})$ is the density function of a normal random vector with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{A}$, $G_m\left(\chi_{m,\alpha}^2; \Delta\right)$ is the density function of a noncentral chi-square distribution with m degrees of freedom and noncentrality parameter Δ,

$$\Delta = \frac{1}{2} \sum_{\substack{i,j \\ i<j}}^I \frac{d_{ij}^2}{p_{ij}} - \sum_{\substack{i,j \\ i<j}} \frac{d_{ij} d_{ji}}{p_{ij}},$$

evaluated in $\chi_{m,\alpha}^2$ and $E[\boldsymbol{\delta}] = \left\{ \mathbf{t} : (\mathbf{t} + \boldsymbol{J}^*(\boldsymbol{\theta})\boldsymbol{\delta})^T (\mathbf{t} + \boldsymbol{J}^*(\boldsymbol{\theta})\boldsymbol{\delta}) \geq \chi_{m,\alpha}^2 \right\}$, where $\boldsymbol{J}^*(\boldsymbol{\theta}) = \mathbf{I} - \boldsymbol{K}(\boldsymbol{\theta})$, being $\boldsymbol{K}(\boldsymbol{\theta}) = \boldsymbol{A}(\boldsymbol{\theta}) \left(\boldsymbol{A}(\boldsymbol{\theta})^T \boldsymbol{A}(\boldsymbol{\theta}) \right)^{-1} \boldsymbol{A}(\boldsymbol{\theta})^T$ and $\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2} \frac{\partial \boldsymbol{p}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. $\boldsymbol{D}_{\boldsymbol{a}}$ denotes the diagonal matrix with vector $\boldsymbol{a}$ in the main diagonal.

§3. Asymptotic bias and asymptotic distributional quadratic risk

Let $\boldsymbol{\theta}=\boldsymbol{g}(\boldsymbol{\beta})$ for some $\boldsymbol{\beta}\in B$. For a suitable estimator $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{*})$ of $\boldsymbol{p}(\boldsymbol{\theta})$ its asymptotic bias, under $H_{1,n}$, is defined by

$$B\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{*})\right)=\lim_{n\to\infty}E\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{*})-\boldsymbol{p}_{n}\right)\right].$$

In the next theorem we shall obtain the asymptotic bias for $\widehat{\boldsymbol{p}}$, $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ and $\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})$.

Theorem 1. *Under $H_{1,n}$ we have,*

$$B\left(\widehat{\boldsymbol{p}}\right)=\mathbf{0},\ B\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)=-\boldsymbol{J}^{*}\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}\ \textit{and}\ B\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)=-\boldsymbol{J}^{*}\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta)$$

Proof. Under $H_{1,n}$, $\lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_{n}\right)\right]=\mathbf{0}$. Therefore, $B\left(\widehat{\boldsymbol{p}}\right)=\mathbf{0}$. Now

$$\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}_{n}\right)=\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)-\boldsymbol{\delta}=\boldsymbol{Z}_{n}-\boldsymbol{\delta}.$$

Therefore, $B\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)=\lim_{n\to\infty}E\left[\boldsymbol{Z}_{n}\right]-\boldsymbol{\delta}=\boldsymbol{K}\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}-\boldsymbol{\delta}=\left(\boldsymbol{K}\left(\boldsymbol{\theta}\right)-\mathbf{I}\right)\boldsymbol{\delta}=-\boldsymbol{J}^{*}\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}$. Finally,

$$\begin{aligned}
B\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)&=\lim_{n\to\infty}E\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)\boldsymbol{I}_{\left(\mathbf{0},\chi^2_{m,\alpha}\right)}(T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right]\\
&+\lim_{n\to\infty}E\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)\boldsymbol{I}_{\left[\mathbf{0},\chi^2_{m,\alpha}\right)}(T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right]-\boldsymbol{\delta}\\
&=\lim_{n\to\infty}E\left[\boldsymbol{Z}_{n}\right]G_m(\chi^2_{m,\alpha};\Delta)+\int\mathbf{x}dF_2(\mathbf{x})-\boldsymbol{\delta}=\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)\\
&+\int\mathbf{x}\int\limits_{E[\boldsymbol{\delta}]}\phi_{N(\mathbf{0},\boldsymbol{K}(\boldsymbol{\theta}))}\left(\mathbf{x}-\boldsymbol{\delta}-\mathbf{t}\right)d\Phi_{N(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}\left(\mathbf{t}\right)-\boldsymbol{\delta}\\
&=\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)+\boldsymbol{\delta}\left(1-G_m(\chi^2_{m,\alpha};\Delta)\right)\\
&+\int\limits_{E[\boldsymbol{\delta}]}\mathbf{t}d\Phi_{N(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}\left(\mathbf{t}\right)-\boldsymbol{\delta},
\end{aligned}$$

being $\boldsymbol{B}\left(\boldsymbol{\theta}\right)=\boldsymbol{I}-\boldsymbol{p}(\boldsymbol{\theta})^{1/2}(\boldsymbol{p}(\boldsymbol{\theta})^{T})^{1/2}-\boldsymbol{K}\left(\boldsymbol{\theta}\right)$.
It is well-known, see formula 5.6 in Sen (1979), that

$$\int\limits_{\mathrm{E}^{*}[\mathbf{a}]}\mathbf{t}d\Phi_{N(\mathbf{0},\boldsymbol{A})}\left(\mathbf{t}\right)=\mathbf{a}\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right]$$

being $E^{*}\left[\mathbf{a}\right]=\left\{\mathbf{t}:\left(\mathbf{t}+\mathbf{a}\right)^{T}\boldsymbol{A}^{-1}\left(\mathbf{t}+\mathbf{a}\right)>\chi^2_{m,\alpha}\right\}$.
In our case it is not possible to apply directly this result because in our case in the set $E\left[\mathbf{a}\right]$, $E\left[\boldsymbol{\delta}\right]$, does not appear matrix $\boldsymbol{A}^{-1}$ but we can overcome this problem in the following way: $\boldsymbol{B}\left(\boldsymbol{\theta}\right)$ is an idempotent matrix with rank $m<I^2$. Therefore there exists an orthogonal matrix $\boldsymbol{C}=(\boldsymbol{C}_1,\boldsymbol{C}_2)$ such that

$$\boldsymbol{C}^{T}\boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}=\begin{pmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}. \tag{12}$$

and

$$\boldsymbol{C}^T\boldsymbol{C}=\mathbf{I}_{I^2}. \tag{13}$$

Based on (12) we have $\boldsymbol{C}_1^T\boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}_1=\mathbf{I}_m$ and based on (13)

$$\left(\begin{array}{cc}\boldsymbol{C}_1^T\boldsymbol{C}_1 & \boldsymbol{C}_1^T\boldsymbol{C}_2\\ \boldsymbol{C}_2^T\boldsymbol{C}_1 & \boldsymbol{C}_2^T\boldsymbol{C}_2\end{array}\right)=\left(\begin{array}{cc}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right).$$

We define the random normal vector $\boldsymbol{X}=\left(X_1,...,X_m\right)^T$ with mean vector $\boldsymbol{C}_1^T\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}$ and variance-covariance matrix $\mathbf{I}_m$. We can write $\boldsymbol{T}=\boldsymbol{C}_1\boldsymbol{X}$ and we have $E\left[\boldsymbol{T}\right]=\boldsymbol{C}_1\boldsymbol{C}_1^T\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}=\boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}=\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}$ (the last equality follows because the matrix $\boldsymbol{B}\left(\boldsymbol{\theta}\right)$ is idempotent and its eigenvalues are 0 or 1), $Var\left[\boldsymbol{T}\right]=\boldsymbol{C}_1^T\boldsymbol{C}_1=\boldsymbol{B}\left(\boldsymbol{\theta}\right)$. The last equality follows because

$$\begin{aligned}\boldsymbol{C}_1^T\boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}_1=\mathbf{I}_m &\Longleftrightarrow \boldsymbol{C}_1\boldsymbol{C}_1^T\boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}_1=\boldsymbol{C}_1\Longleftrightarrow \boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}_1=\boldsymbol{C}_1\\ &\Longleftrightarrow \boldsymbol{B}\left(\boldsymbol{\theta}\right)\boldsymbol{C}_1\boldsymbol{C}_1^T=\boldsymbol{C}_1\boldsymbol{C}_1^T\Longleftrightarrow \boldsymbol{B}\left(\boldsymbol{\theta}\right)=\boldsymbol{C}_1\boldsymbol{C}_1^T.\end{aligned}$$

We can also observe that $\boldsymbol{T}^T\boldsymbol{T}=\boldsymbol{X}^T\boldsymbol{C}_1^T\boldsymbol{C}_1\boldsymbol{X}=\boldsymbol{X}^T\boldsymbol{X}$. Now we can write

$$\begin{aligned}\int\limits_{\mathrm{E}[\boldsymbol{\delta}]}\mathbf{t}d\Phi_{N(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}\left(\mathbf{t}\right)&=\int\limits_{\mathrm{E}[\mathbf{0}]}\mathbf{t}d\Phi_{N(\mathbf{J}^*(\boldsymbol{\theta})\boldsymbol{\delta},\boldsymbol{B}(\boldsymbol{\theta}))}\left(\mathbf{t}\right)=\boldsymbol{C}_1\int\limits_{\left\{\mathbf{x}:\mathbf{x}^T\mathbf{x}>\chi^2_{m,\alpha}\right\}}\mathbf{x}d\Phi_{\mathcal{N}(\mathbf{C}_1^T\mathbf{J}^*(\boldsymbol{\theta})\boldsymbol{\delta},\mathbf{I})}\left(\mathbf{x}\right)\\ &=\boldsymbol{C}_1\int\limits_{\left\{\mathbf{x}:\left(\mathbf{x}+\mathbf{C}_1^T\mathbf{J}^*(\boldsymbol{\theta})\boldsymbol{\delta}\right)^T\left(\mathbf{x}+\mathbf{C}_1^T\mathbf{J}^*(\boldsymbol{\theta})\boldsymbol{\delta}\right)>\chi^2_{m,\alpha}\right\}}\mathbf{x}d\Phi_{N(\mathbf{0},\mathbf{I})}\left(\mathbf{x}\right)\\ &=\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}B\left(\boldsymbol{p}^{\mathrm{pre}}_{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right)&=\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)+\boldsymbol{\delta}I-\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)\\ &+\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)-\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta)-\boldsymbol{\delta}-\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta)\\ &=\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta).\end{aligned}$$

□

Let $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)$ be a suitable estimator of $\boldsymbol{p}(\boldsymbol{\theta})$ and $\boldsymbol{W}$ a given semi-definite matrix. The standard loss function is

$$L\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*),\boldsymbol{p}(\boldsymbol{\theta})\right)=\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)-\boldsymbol{p}(\boldsymbol{\theta})\right)^T\boldsymbol{W}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)-\boldsymbol{p}(\boldsymbol{\theta})\right)$$

and the asymptotic distributional quadratic risk (ADQR) of $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)$, under $H_{1,n}$, is given by

$$R\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*);\boldsymbol{W}\right)=\lim_{n\to\infty}E\left[\sqrt{n}\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta})}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)-\boldsymbol{p}_n\right)^T\boldsymbol{W}\sqrt{n}\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta})}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^*)-\boldsymbol{p}_n\right)\right].$$

We have the following result:

Theorem 2. *Let $\boldsymbol{\theta}=\boldsymbol{g}(\boldsymbol{\beta})$ for some $\boldsymbol{\beta}\in B$. Under $H_{1,n}$ we have,*

$$\begin{aligned}
R\left(\widehat{\boldsymbol{p}};\boldsymbol{W}\right) &= \operatorname{trace}\left(\boldsymbol{J}(\boldsymbol{\theta})\right)\\
R\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= \operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)+\boldsymbol{\delta}^T\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)^T\boldsymbol{W}\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}\\
R\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= \operatorname{trace}\left(\boldsymbol{W}\boldsymbol{J}(\boldsymbol{\theta})\right)-\operatorname{trace}\left(\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{W}\right)G_{m+2}(\chi^2_{m,\alpha};\Delta)\\
&\quad-\boldsymbol{\delta}^T\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)^T\boldsymbol{W}\boldsymbol{J}^*\left(\boldsymbol{\theta}\right)\boldsymbol{\delta}\left\{-2G_{m+2}(\chi^2_{m,\alpha};\Delta)+G_{m+4}(\chi^2_{m,\alpha};\Delta)\right\}.
\end{aligned}$$

Proof. We shall use in the proof of this theorem the following well-known fact,

$$E\left[\left(\boldsymbol{Y}-\mathbf{a}\right)^T\boldsymbol{A}\left(\boldsymbol{Y}-\mathbf{a}\right)\right]=\operatorname{trace}\left(\boldsymbol{A}\boldsymbol{\Sigma}\right)+\left(\boldsymbol{\mu}-\mathbf{a}\right)^T\boldsymbol{A}\left(\boldsymbol{\mu}-\mathbf{a}\right)$$

for a normal random vector $\boldsymbol{Y}$ with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. We know that $\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}(\widehat{\boldsymbol{p}}-\boldsymbol{p}_n)\underset{n\to\infty}{\overset{L}{\to}}\mathcal{N}(\mathbf{0},\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{\mathbf{-1/2}}\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2})$ and $\sqrt{n}(\widehat{\boldsymbol{p}}-\boldsymbol{p}_n)\underset{n\to\infty}{\overset{L}{\to}}\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})})$, but $\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{\mathbf{-1/2}}\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta})}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}=\boldsymbol{J}(\boldsymbol{\theta})$. Therefore,

$$R\left(\widehat{\boldsymbol{p}};\boldsymbol{W}\right)=\lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_n\right)^T\boldsymbol{W}\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_n\right)\right)\right]=\operatorname{trace}\left(\boldsymbol{J}(\boldsymbol{\theta})\right),$$

and

$$\begin{aligned}
R\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= \lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)^T\boldsymbol{W}\left(\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)\right]\\
&\quad-\lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)^T\right]\boldsymbol{W}\boldsymbol{\delta}\\
&\quad-\boldsymbol{\delta}^T\boldsymbol{W}\lim_{n\to\infty}E\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)\right]+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}\\
&= \operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)+\boldsymbol{\delta}^T\boldsymbol{K}(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}-\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}\\
&\quad-\boldsymbol{\delta}^T\boldsymbol{K}(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{\delta}+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}\\
&= \operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)+\boldsymbol{\delta}^T\left(\boldsymbol{K}(\boldsymbol{\theta})-\boldsymbol{I}\right)^T\boldsymbol{W}\left(\boldsymbol{K}(\boldsymbol{\theta})-\boldsymbol{I}\right)\boldsymbol{\delta}.
\end{aligned}$$

In relation to $\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ we have,

$$\begin{aligned}
&R\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right)=\lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}_n\right)^T\boldsymbol{W}\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\right.\\
&\left.\times\,\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}_n\right]-\lim_{n\to\infty}\mathrm{E}\left[\left(\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)-\boldsymbol{\delta}\right)^T\right]\boldsymbol{W}\boldsymbol{\delta}\\
&-\boldsymbol{\delta}^T\boldsymbol{W}\lim_{n\to\infty}\mathrm{E}\left[\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})-\boldsymbol{p}(\boldsymbol{\theta})\right)\right]+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}=l_1+l_2+l_3+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}.
\end{aligned}$$

Now we are going to get l_1, l_2 and l_3. In relation to l_1 we have,

$$l_1=\int\mathbf{x}^T\boldsymbol{W}\mathbf{x}f_{\boldsymbol{W}}\left(\mathbf{x}\right)d\mathbf{x}=\int\mathbf{x}^T\boldsymbol{W}\mathbf{x}dF_1\left(\mathbf{x}\right)+\int\mathbf{x}^T\boldsymbol{W}\mathbf{x}dF_2\left(\mathbf{x}\right)$$

but

$$\begin{aligned}\int \mathbf{x}^T \boldsymbol{W}\mathbf{x} dF_1(\mathbf{x}) &= G_m(\chi^2_{m,\alpha};\Delta)\int \mathbf{x}^T \boldsymbol{W}\mathbf{x}\phi_{\mathcal{N}(\mathbf{0},\boldsymbol{K}(\boldsymbol{\theta}))}(\mathbf{x}-\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta})d\mathbf{x} \\ &= G_m(\chi^2_{m,\alpha};\Delta)\left(\operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})+(\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta})^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}\right)\right.\end{aligned}$$

On the other hand

$$\begin{aligned}\int \mathbf{x}^T \boldsymbol{W}\mathbf{x} dF_2(\mathbf{x}) &= \operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)\left(1-G_m\left(\chi^2_{m,\alpha};\Delta\right)\right)+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}\left(1-G_m\left(\chi^2_{m,\alpha};\Delta\right)\right) \\ &\quad + b_1+b_2+b_3.\end{aligned}$$

But

$$\begin{aligned}b_1 &= \int\limits_{E[\boldsymbol{\delta}]} \mathbf{t}^T\boldsymbol{W}\boldsymbol{\delta}\phi_{\mathcal{N}(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}(\mathbf{t})\,\mathbf{dt} = \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right]\boldsymbol{W}\boldsymbol{\delta} \\ b_2 &= \int\limits_{\mathbf{E}[\boldsymbol{\delta}]} \boldsymbol{\delta}^T\boldsymbol{W}\mathbf{t}\phi_{\mathcal{N}(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}(\mathbf{t})\,\mathbf{dt} = \boldsymbol{\delta}^T\boldsymbol{W}\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right]\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}.\end{aligned}$$

In relation to b_3 we have,

$$\begin{aligned}b_3 &= \int\limits_{\mathbf{E}[\boldsymbol{\delta}]} \mathbf{t}^T\boldsymbol{W}\mathbf{t}\phi_{\mathcal{N}(\mathbf{0},\boldsymbol{B}(\boldsymbol{\theta}))}(\mathbf{t})\,d\mathbf{t} = \int\limits_{\mathbf{E}[\mathbf{0}]} \mathbf{t}^T\boldsymbol{W}\mathbf{t}\phi_{\mathcal{N}(\boldsymbol{J}^*(\boldsymbol{\theta})\boldsymbol{\delta},\boldsymbol{B}(\boldsymbol{\theta}))}(\mathbf{t})\,d\mathbf{t} \\ &= \int\limits_{\mathbf{E}[\mathbf{0}]} \mathbf{x}^T\boldsymbol{C}_1^T\boldsymbol{W}\boldsymbol{C}_1\mathbf{x}\phi_{\mathcal{N}(\boldsymbol{C}_1^T\boldsymbol{J}^*(\boldsymbol{\theta})\boldsymbol{\delta},\mathbf{I})}(\mathbf{x})\,d\mathbf{x} = \int\limits_{\mathbf{E}^*[\boldsymbol{\delta}]} \mathbf{x}^T\boldsymbol{C}_1^T\boldsymbol{W}\boldsymbol{C}_1\mathbf{x}\phi_{\mathcal{N}(\mathbf{0},\mathbf{I})}(\mathbf{x})\,d\mathbf{x},\end{aligned}$$

being $\mathrm{E}^*[\boldsymbol{\delta}] = \left\{\mathbf{t}:\left(\mathbf{t}+\boldsymbol{C}_1^T\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\right)^T\left(\mathbf{t}+\boldsymbol{C}_1^T\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\right) > \chi^2_{m,\alpha}\right\}$. Therefore

$$\begin{aligned}b_3 &= \left(1-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right)\operatorname{trace}\left(\boldsymbol{C}_1^T\boldsymbol{W}\boldsymbol{C}_1\right)-\boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{C}_1\boldsymbol{C}_1^T\boldsymbol{W}\boldsymbol{C}_1\boldsymbol{C}_1^T\boldsymbol{\delta}\boldsymbol{J}^*(\boldsymbol{\theta}) \\ &\quad\times\left(G_m(\chi^2_{m,\alpha};\Delta)-2G_{m+2}(\chi^2_{m,\alpha};\Delta)+G_{m+4}(\chi^2_{m,\alpha};\Delta)\right)\end{aligned}$$

We can conclude that

$$\begin{aligned}l_1 &= G_m(\chi^2_{m,\alpha};\Delta)\left(\operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})+\boldsymbol{\delta}^T\boldsymbol{K}(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}\right)\right. \\ &\quad+\operatorname{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)\left(1-G_m\left(\chi^2_{m,\alpha};\Delta\right)\right)+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}\left(1-G_m\left(\chi^2_{m,\alpha};\Delta\right)\right) \\ &\quad+\boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{\delta}\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right] \\ &\quad+\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\left[G_m(\chi^2_{m,\alpha};\Delta)-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right] \\ &\quad+\left(1-G_{m+2}(\chi^2_{m,\alpha};\Delta)\right)\operatorname{trace}\left(\boldsymbol{B}(\boldsymbol{\theta})\,\boldsymbol{W}\right)-\boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta} \\ &\quad\times\left(G_m(\chi^2_{m,\alpha};\Delta)-2G_{m+2}(\chi^2_{m,\alpha};\Delta)+G_{m+4}(\chi^2_{m,\alpha};\Delta)\right)\end{aligned}$$

Now it is not very difficult to see l_2 is given by

$$l_2 = \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{W}\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta)-\boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}.$$

In relation to l_3 we have,

$$l_3 = -\boldsymbol{\delta}^T \boldsymbol{W} \lim_{n\to\infty} E\sqrt{n}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta})}^{-1/2}\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2}) - \boldsymbol{p}(\boldsymbol{\theta})\right) = -\boldsymbol{\delta}^T \boldsymbol{W}\left(B\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2})\right) + \boldsymbol{\delta}\right)$$
$$= \boldsymbol{\delta}^T \boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta} G_{m+2}(\chi^2_{m,\alpha};\Delta) - \boldsymbol{\delta}^T \boldsymbol{W}\boldsymbol{\delta}.$$

Finally we have,

$$\begin{aligned} R\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= G_m(\chi^2_{m,\alpha};\Delta)\left(\mathrm{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta}) + (\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta})^T \boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}\right)\right. \\ &+ \mathrm{trace}\left(\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\right)\left(1 - G_m\left(\chi^2_{m,\alpha};\Delta\right)\right) + \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}\left(1 - G_m\left(\chi^2_{m,\alpha};\Delta\right)\right) \\ &+ \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{\delta}\left[G_m(\chi^2_{m,\alpha};\Delta) - G_{m+2}(\chi^2_{m,\alpha};\Delta)\right] \\ &+ \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\left[G_m(\chi^2_{m,\alpha};\Delta) - G_{m+2}(\chi^2_{m,\alpha};\Delta)\right] \\ &- \left(1 - G_{m+2}(\chi^2_{m,\alpha};\Delta)\right)\mathrm{trace}\left(\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{W}\right) \\ &- \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\left\{G_m(\chi^2_{m,\alpha};\Delta) - 2G_{m+2}(\chi^2_{m,\alpha};\Delta)\right. \\ &\left.+G_{m+4}(\chi^2_{m,\alpha};\Delta)\right\} + \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{W}\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta) - \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta} \\ &+ \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}G_{m+2}(\chi^2_{m,\alpha};\Delta) - \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}. \end{aligned}$$

Simplifying, we get

$$\begin{aligned} R\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= \mathrm{trace}\left(\boldsymbol{W}\boldsymbol{J}(\boldsymbol{\theta})\right) - \mathrm{trace}\left(\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{W}\right)G_{m+2}(\chi^2_{m,\alpha};\Delta) \\ &+ G_m(\chi^2_{m,\alpha};\Delta)\boldsymbol{\delta}^T\boldsymbol{K}(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta} - \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta) \\ &+ \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta) - \boldsymbol{\delta}^T\boldsymbol{K}(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta) \\ &+ \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta) - \boldsymbol{\delta}^T\boldsymbol{W}\boldsymbol{K}(\boldsymbol{\theta})\boldsymbol{\delta}G_m(\chi^2_{m,\alpha};\Delta) \\ &- \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\left\{G_m(\chi^2_{m,\alpha};\Delta) - 2G_{m+2}(\chi^2_{m,\alpha};\Delta)\right. \\ &\left.+G_{m+4}(\chi^2_{m,\alpha};\Delta)\right\}. \end{aligned}$$

Finally,

$$\begin{aligned} R\left(\boldsymbol{p}_{\phi_1}^{\mathrm{pre}}(\widehat{\boldsymbol{\theta}}^{\phi_2});\boldsymbol{W}\right) &= \mathrm{trace}\left(\boldsymbol{W}\boldsymbol{J}(\boldsymbol{\theta})\right) - \mathrm{trace}\left(\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{W}\right)G_{m+2}(\chi^2_{m,\alpha};\Delta) \\ &- \boldsymbol{\delta}^T\boldsymbol{J}^*(\boldsymbol{\theta})^T\boldsymbol{W}\boldsymbol{J}^*(\boldsymbol{\theta})\,\boldsymbol{\delta}\left\{-2G_{m+2}(\chi^2_{m,\alpha};\Delta) + G_{m+4}(\chi^2_{m,\alpha};\Delta)\right\}. \end{aligned}$$

□

Acknowledgements

This work was partially supported by Grants MTM2009-10072 and BSCH-UCM2008-910707.

References

[1] ALI, S. M., AND SILVEY, S. D. A General Class of Coefficients of Divergence of one Distribution from another. *Journal of the Royal Statistical Society, Series B*, 28 (1966), 131–142.

[2] BANCROFT, T. A. On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics*, 15 (1944), 190–204.

[3] BOWKER, A. A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43 (1948), 572–574.

[4] MENÉNDEZ, M. L., PARDO, L., AND PARDO, M. C. Preliminary test estimators and phi-divergence measures in generalized linear models with binary data. *Journal of Multivariate Analysis*, 99 (2008), 2265–2284.

[5] CSISZÀR, I. Eine Informations theorestiche Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the mathematical Institute of Hungarian Academy of Sciences, Series A*, 8 (1963), 84–108.

[6] MORALES, D. PARDO, L., AND VAJDA, I. Asymptotic divergences of estimates of discrete distributions. *Journal Statistical Planning and Inference*, 48 (1995), 347–369.

[7] PARDO, L. *Statistical Inference Based on Divergence Measures*. Chapman & Hall, CRC, New York, 1996.

[8] PARDO, L. AND MENÉNDEZ, M. L. On some pre-test and Stein-rule phi-divergence test estimators in the independence model of categorical data. *Journal of Statistical Planning and Inference*, 138 (2007), 2163-2179.

[9] SALEH, A. K. MD. E. *Theory of Preliminary Test and Stein-Type Estimation with Applications*. Wiley, New York, 2006.

[10] SEN, P. K. Asymptotic properties of maximum likelihood estimators based on conditional specifications. *Annals of Statistic*, 7 (1979), 1019-1033.

Leandro Pardo
Departamento de Estadística e I. O.
Universidad Complutense de Madrid
Plaza de Ciencias 3, 28040 Madrid, Spain
lpardo@mat.ucm.es

Nirian Martín
Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid 126, 28903 Getafe, Spain
nirian.martin@uc3m.es

Monografías Matemáticas García de Galdeano **36**, 137–145 (2010)

OPTIMAL LOT SIZE FOR AN INVENTORY SYSTEM WITH A STEP FUNCTION OF THE CUSTOMERS' WAITING TIME

Luis A. San-José, Joaquín Sicilia and Juan García-Laguna

Abstract. In this work we consider a continuous review inventory control system where demand on the stockout period is partially backlogged. More specifically, the backlogged demand ratio is a two piece function of time the customers have to wait up to receiving the next replenishment. This ratio is a known constant when the waiting time is less than the maximum time the customers are willing to wait. Otherwise, the ratio vanishes when the waiting time exceeds that maximum. Moreover, we suppose that both backorder unit cost and lost sale cost are made up of a fixed cost and a variable cost which depends on the length of the shortage time. A general procedure to determine the optimal policy and the minimum inventory cost is developed. This model generalizes several inventory systems analyzed by different authors. Numerical examples are used to illustrate the theoretical results.

Keywords: inventory models, partial backlogging, stockout costs.

AMS classification: AMS 90B05.

§1. Introduction

In any inventory system it can happen that a part of the demand may not be satisfied with the current stock. In this event, we speak about the system is out of stock. Frequently, when there are shortages, the classical inventory models suppose that either all customers wait until the arrival of the next order (complete backorder case) or all customers leave the system (lost sales case). However, in many real systems, some customers are able to wait for the next order to satisfy their demands during the stockout period, while others do not wish to or cannot wait and they have to fill their demands from other sources. This situation is modeled by the consideration of partial backordering in the formulation of the mathematical models.
A common characteristic of the models with partial backlogging is to assume a fixed penalty cost per lost unit. However, in practice, the customers usually make decision to wait until the next replenishment or not, according to the time they would have to wait and the possible compensation from the firm if they wait. Consequently, in both cases (backordered or lost demand), the commercial prestige depends on the time that elapses until the arrival of the next replenishment. This point of view is assumed in Chern, Chan & Teng (2005) and San-José, Sicilia & García-Laguna (2009).
In this work, we propose and study an inventory model where the customers are willing to wait for their orders to be filled at most a fixed time. This situation is cited in Lee and

Nahmias (1993, p. 8). Moreover, we will consider that both backorder unit cost and lost sale cost are made up of a fixed cost and a variable cost which depends on the length of the shortage time. As a result, the proposed model is a general framework that includes several previous models.

§2. Assumptions and notation

The mathematical model discussed in this work is based on the following assumptions:

1. The item is a single product with independent demand.
2. The planning horizon is infinite.
3. The replenishment rate is infinite.
4. The inventory is continuously revised.
5. The demand rate is known and constant.
6. The ordering cost per order is known and constant.
7. The unit purchasing cost and the selling price per unit are constant.
8. The holding cost is a linear function based on average inventory.
9. The model allows shortages, which are partially backlogged.
10. The fraction of backlogged demand is described by a function, which depends on the amount of time a customer waits before receiving the item.
11. The cost of a backorder includes a fixed cost and a cost which is proportional to the length of time for which backorder exists.
12. The goodwill cost of a lost sale includes a fixed cost and a cost which is proportional to the length of time for which lost sales exist.

We adopt the following notation for the model to be studied:

Input parameters

D demand per unit time (> 0).
K ordering cost per order (> 0).
p unit purchasing cost (> 0).
s unit selling price $(s > p)$.
h holding cost per unit per unit of time (> 0).
ω_o fixed backorder cost per unit, independent of time (≥ 0).
ω backorder cost per unit and per unit time (≥ 0).
π_o fixed goodwill cost per lost unit, independent of time, that is, the fixed cost derived of a unit lost sale excluding the loss of profit (≥ 0).
π goodwill cost per lost unit and per unit time (≥ 0).
ρ maximum fraction of backordered demand $(0 \leq \rho \leq 1)$.

Additional notation

$\widetilde{\pi}$ fixed cost of a lost sale, including the profit loss and the fixed goodwill cost, that is, $\widetilde{\pi} = \pi_o + s - p$.
ξ_o fixed unit shortage cost including the loss of profit, $\xi_o = \rho\omega_o + (1-\rho)\widetilde{\pi}$.
ξ average shortage cost dependent on time, $\xi = \rho\omega + (1-\rho)\pi$.
$I(t)$ net stock (on hand – backorders) level at time t.
τ amount of time the customers wait before receiving the good during the stockout period (> 0).
q order quantity or lot size per cycle (≥ 0).
b demand during the stockout period (≥ 0).
$\beta(\tau)$ fraction of demand which is backlogged.

Decision variables

T length of the inventory cycle (> 0).
Ψ length of the inventory cycle over which the net stock level is less than or equal to zero (≥ 0).

§3. The model

Next, we analyze an inventory model with partial backlogging, where the backlogged demand rate is a step function of the customers' waiting time. Moreover, we suppose that the cumulative unit stockout cost (backordered cost and lost sale cost) is a linear function of the length of time for which the shortage exists. We also consider that the fraction of backlogged demand, which represents the behavior of the customers faces with stockout, is given by

$$\beta(\tau) = \begin{cases} \rho & \text{if} \quad 0 < \tau \leq a \\ 0 & \text{if} \quad \tau > a \end{cases}, \text{ with } 0 \leq \rho \leq 1 \text{ and } a \geq 0,$$

being ρ the maximum fraction of backordered demand and a the maximum waiting time that the customers are willing to wait up to receiving the product.
Taking into account the above assumptions, the net inventory level during the stockout period can be expressed as

$$I(t) = \begin{cases} -D\rho(\Psi - T + t) & \text{if} \quad a \geq \Psi \\ 0 & \text{if} \quad a < \Psi \quad \text{and} \quad T - \Psi \leq t \leq T - a \\ -D\rho(a - T + t) & \text{if} \quad a < \Psi \quad \text{and} \quad T - a < t < T \end{cases} \tag{1}$$

Evidently, the objective is to maximize the profit per unit time. The significant amounts at each cycle are the following:

- Total revenue: $s[D(T-\Psi) - I_T]$.
- Purchasing cost: $p[D(T-\Psi) - I_T]$.
- Ordering cost: K.
- Total holding cost: $h\int_0^{T-\Psi} I(t)dt = \frac{Dh}{2}(T-\Psi)^2$.

- Total backordering cost: $D\omega_o \int_0^{\Psi} \beta(\tau)dr + \int_{T-\Psi}^{T} \omega[-I(t)]dt =$
$$= \begin{cases} D\rho\omega_o\Psi + D\rho\omega\Psi^2/2 & \text{if} \quad 0 \leq \Psi \leq a \\ D\rho\omega_o a + D\rho\omega a^2/2 & \text{if} \quad \Psi > a \end{cases}.$$

- Total lost sale cost: $D\pi_o \int_0^{\Psi} [1-\beta(\tau)]d\tau + \pi D \int_0^{\Psi} \tau[1-\beta(\tau)]d\tau =$
$$= \begin{cases} D(1-\rho)\pi_o\Psi + D(1-\rho)\pi\Psi^2/2 & \text{if} \quad 0 \leq \Psi \leq a \\ D\pi_o(\Psi - \rho a) + D\pi(\Psi^2 - \rho a^2)/2 & \text{if} \quad \Psi > a \end{cases}.$$

Therefore, the gain or profit per unit time can be written as

$$G(T,\Psi) = (s-p)D - C(T,\Psi), \tag{2}$$

where

$$C(T,\Psi) = \begin{cases} C_1(T,\Psi) & \text{if} \quad 0 \leq \Psi \leq a \\ C_2(T,\Psi) & \text{if} \quad \Psi > a \end{cases},$$

with

$$C_1(T,\Psi) = \frac{1}{T}\left[K + \tfrac{Dh}{2}(T-\Psi)^2 + D\xi_o\Psi + \tfrac{D\xi}{2}\Psi^2\right]$$

and

$$C_2(T,\Psi) = \tfrac{1}{T}\left[K + \tfrac{Dh}{2}(T-\Psi)^2 + D(\xi_o - \widetilde{\pi})a + \tfrac{D(\xi-\pi)a^2}{2} + \widetilde{\pi}D\Psi + \tfrac{D\pi}{2}\Psi^2\right].$$

Hence, the problem consists of determining the decision variables T and Ψ, with $T > 0$, $\Psi \geq 0$ and $\Psi \leq T$, such that the function $G(T,\Psi)$ given in (2) is maximized.

Note. The lot size q and the demand b during the stockout period are given by

$$q = \begin{cases} D\left[T - (1-\rho)\Psi\right] & \text{if} \quad 0 \leq \Psi \leq a \\ D\left[T - \Psi + \rho a\right] & \text{if} \quad \Psi > a \end{cases} \quad \text{and} \quad b = D\Psi.$$

Thus, if we calculate the optimal policy (T^*, Ψ^*), we will specific the lot size q and the reorder point.

§4. Solution of the model

After some algebraic manipulations, the functions $C_1(T,\Psi)$ and $C_2(T,\Psi)$ also can be expressed as

$$C_1(T,\Psi) = \tfrac{K + D\xi_o\Psi + D(\xi+h)\Psi^2/2}{T} + \tfrac{Dh}{2}T - Dh\Psi$$

and

$$C_2(T,\Psi) = \tfrac{K + D(\xi_o - \widetilde{\pi})a + D(\xi-\pi)a^2/2 + \widetilde{\pi}D\Psi + D(\pi+h)\Psi^2/2}{T} + \tfrac{Dh}{2}T - Dh\Psi.$$

Therefore, for a fixed value of Ψ, the function $C(T,\Psi)$ is strictly convex. In consequence, it attains its minimum at the point

$$T^*(\Psi) = \begin{cases} \sqrt{\frac{2K+2D\xi_o\Psi+D(\xi+h)\Psi^2}{Dh}} & \text{if} \quad 0 \le \Psi \le a \\ \sqrt{\frac{2K+2D(\xi_o-\widetilde{\pi})a+D(\xi-\pi)a^2+2\widetilde{\pi}D\Psi+D(h+\pi)\Psi^2}{Dh}} & \text{if} \quad \Psi > a \end{cases}$$

with optimal value

$$C(T^*(\Psi), \Psi) = Dh\left[T^*(\Psi) - \Psi\right]. \tag{3}$$

Note that $C(T^*(0), 0) = \sqrt{2KDh}$, $C(T^*(\infty), \infty) = \widetilde{\pi}D$ if $\pi = 0$, and $C(T^*(\infty), \infty) = \infty$ for all $\pi > 0$. Moreover, the first derivative of the function $C(T^*(\Psi), \Psi)$ is

$$\frac{dC(T^*(\Psi), \Psi)}{d\Psi} = \frac{L(\Psi)}{\alpha(\Psi)T^*(\Psi)}, \tag{4}$$

where

$$\alpha(\Psi) = \begin{cases} D\xi_o + D(h+\xi)\Psi + DhT^*(\Psi) & \text{if} \quad 0 \le \Psi < a \\ D\widetilde{\pi} + D(h+\pi)\Psi + DhT^*(\Psi) & \text{if} \quad \Psi > a \end{cases}, \tag{5}$$

and

$$L(\Psi) = \begin{cases} L_1(\Psi) & \text{if} \quad 0 \le \Psi < a \\ L_2(\Psi) & \text{if} \quad \Psi > a \end{cases}, \tag{6}$$

with

$$\begin{aligned} L_1(\Psi) &= (\xi_o D)^2 - 2KhD + 2D^2\xi_o\xi\Psi + D^2\xi(h+\xi)\Psi^2, \\ L_2(\Psi) &= \theta + 2D^2[\widetilde{\pi} + (h+\pi)a]\pi(\Psi - a) + D^2\pi(h+\pi)(\Psi - a)^2 \end{aligned}$$

and

$$\theta = (D\widetilde{\pi})^2 - 2KDh + 2D^2[\widetilde{\pi}(h+\pi) - h\xi_o]a + D^2(\pi^2 + 2h\pi - h\xi)a^2.$$

Since $T^*(\Psi) > 0$ and $\alpha(\Psi) > 0$ for all $\Psi \ge 0$, we have

$$\text{sign}\ \frac{dC(T^*(\Psi), \Psi)}{d\Psi} = \text{sign}\ L(\Psi). \tag{7}$$

Thus, we can determine the optimal value of Ψ by studying the function $L(\Psi)$. According to this approach, the following theorem provides a criterion to determine the optimal policy.

Theorem 1. *Let* $\theta = (D\widetilde{\pi})^2 - 2KDh + 2D^2[\widetilde{\pi}(h+\pi) - h\xi_o]a + D^2(\pi^2 + 2h\pi - h\xi)a^2$; $\Delta = (\xi_o D)^2 - 2KDh$; $L_1(a) = \Delta + 2D^2\xi_o\xi a + D^2\xi(h+\xi)a^2$; $C_o = \sqrt{2KDh}$; $C_1 = D\left(\xi\sqrt{\frac{2Kh}{D\xi(h+\xi)} - \frac{h\xi_o^2}{\xi(h+\xi)^2}} + \frac{\xi_o h}{h+\xi}\right)$; $C_2 = D\left(\sqrt{\pi^2(a + \frac{\widetilde{\pi}}{h+\pi})^2 - \frac{\theta\pi}{D^2(h+\pi)}} + \frac{h\widetilde{\pi}}{h+\pi}\right)$.

The optimal solution (T^*, Ψ^*) *which maximizes* $G(T, \Psi)$ *is given in Table 1, where the values of the decision variables are:*

$\Psi_1 = \sqrt{\frac{2Kh}{D\xi(h+\xi)} - \frac{h\xi_o^2}{\xi(h+\xi)^2}} - \frac{\xi_o}{h+\xi}$; $\Psi_2 = \sqrt{(a + \frac{\widetilde{\pi}}{h+\pi})^2 - \frac{\theta}{D^2\pi(h+\pi)}} - \frac{\widetilde{\pi}}{h+\pi}$;

$T_o = \sqrt{2K/Dh}$; $T_a = \sqrt{(2K + 2D\xi_o a + D(h+\xi)a^2)/Dh}$;

$T_1 = \Psi_1 + (\xi_o + \xi\Psi_1)/h$ *and* $T_2 = \Psi_2 + (\widetilde{\pi} + \pi\Psi_2)/h$.

Let us mention one important consequence of the previous theorem.

Corollary 2. *With the notation used in Theorem 1, the maximum total inventory profit is shown in Table 2.*

			$\pi > 0$		$\pi = 0$	
			$\xi > 0$	$\xi = 0$	$\xi > 0$	$\xi = 0$
$\Delta > 0$	$\theta \geq 0$		$(T_o, 0)$	$(T_o, 0)$	$(T_o, 0)$	$(T_o, 0)$
	$\theta < 0$	$C_2 \geq C_o$	$(T_o, 0)$	$(T_o, 0)$	$(T_o, 0)$	$(T_o, 0)$
		$C_2 < C_o$	(T_2, Ψ_2)	(T_2, Ψ_2)	(∞, ∞)	(∞, ∞)
$\Delta = 0$	$\theta > 0$		$(T_o, 0)$	$(T_o + \Psi, \Psi)^{(1)}$	$(T_o, 0)$	$(T_o + \Psi, \Psi)^{(1)}$
	$\theta = 0$		$(T_o, 0)$	$(T_o + \Psi, \Psi)^{(1)}$	$(T_o, 0)$	$(T_o + \Psi, \Psi)^{(2)}$
	$\theta < 0$	$C_2 \geq C_o$	$(T_o, 0)$	——	$(T_o, 0)$	——
		$C_2 < C_o$	(T_2, Ψ_2)	(T_2, Ψ_2)	(∞, ∞)	(∞, ∞)
$\Delta < 0$	$L_1(a) \geq 0$	$\theta \geq 0$	(T_1, Ψ_1)	——	(T_1, Ψ_1)	——
		$\theta < 0$ $C_2 \geq C_1$	(T_1, Ψ_1)	——	(T_1, Ψ_1)	——
		$\theta < 0$ $C_2 < C_1$	(T_2, Ψ_2)	——	(∞, ∞)	——
	$L_1(a) < 0$	$\theta > 0$	(T_a, a)	(T_a, a)	(T_a, a)	(T_a, a)
		$\theta = 0$	(T_a, a)	(T_a, a)	$(C_2/Dh + \Psi, \Psi)^{(3)}$	$(C_2/Dh + \Psi, \Psi)^{(3)}$
		$\theta < 0$	(T_2, Ψ_2)	(T_2, Ψ_2)	(∞, ∞)	(∞, ∞)

(1) for all $\Psi \in [0, a]$

(2) for all $\Psi \in [0, \infty)$

(3) for all $\Psi \in [a, \infty)$

Table 1: Optimal policy (T^*, Ψ^*)

			$\pi > 0$	$\pi = 0$
$\Delta > 0$	$\theta \geq 0$		$(s-p) - \sqrt{2KDh}$	$(s-p)D - \sqrt{2KDh}$
	$\theta < 0$	$C_2 \geq C_o$	$(s-p)D - \sqrt{2KDh}$	$(s-p)D - \sqrt{2KDh}$
		$C_2 < C_o$	$(s-p)D - C_2$	$(s-p-\widetilde{\pi})D$
$\Delta = 0$	$\theta > 0$		$(s-p)D - \sqrt{2KDh}$	$(s-p)D - \sqrt{2KDh}$
	$\theta = 0$		$(s-p)D - \sqrt{2KDh}$	$(s-p)D - \sqrt{2KDh}$
	$\theta < 0$	$C_2 \geq C_o$	$(s-p)D - \sqrt{2KDh}$	$(s-p)D - \sqrt{2KDh}$
		$C_2 < C_o$	$(s-p)D - C_2$	$(s-p-\widetilde{\pi})D$
$\Delta < 0$	$L_1(a) \geq 0$	$\theta \geq 0$	$(s-p)D - C_1$	$(s-p)D - C_1$
		$\theta < 0$, $C_2 \geq C_1$	$(s-p)D - C_1$	$(s-p)D - C_1$
		$\theta < 0$, $C_2 < C_1$	$(s-p)D - C_2$	$(s-p-\widetilde{\pi})D$
	$L_1(a) < 0$	$\theta > 0$	$D[s-p+h(T_a-a)]$	$D[s-p+h(T_a-a)]$
		$\theta = 0$	$D[s-p+h(T_a-a)]$	$(s-p-\widetilde{\pi})D$
		$\theta < 0$	$(s-p)D - C_2$	$(s-p-\widetilde{\pi})D$

Table 2: Maximum inventory profit

§5. Numerical examples

In this section, we illustrate with some examples the solution procedure developed in the previous section.

Example 1 We consider an inventory system which verifies the assumptions described in Section 2. The input data and parameters are $D = 25$, $K = 50$, $h = 0.5$, $\omega_o = 1$, $\omega = 0$, $\pi_o = 0$, $\pi = 1$, $p = 9$, $s = 12$, $\rho = 0.9$ and $a = 2$. Following the development given in the previous section, we have: $\Delta = -350$, $\xi = 0.1$, $L_1(a) = 100$ and $\theta = 19000$. Therefore, applying Theorem 1, the function $G(T, \Psi)$ attains its maximum at the point (T_1, Ψ_1), where $\Psi_1 = 1.65148$ and $T_1 = 4.38178$. Moreover, from Corollary 2, we obtain $G(T^*, \Psi^*) = 40.87129$.

Example 2 We consider the same input data and parameters of Example 1, but changing the value of the maximum waiting time to $a = 1.5$. Now $L_1(a) = -40.625$ and $\theta = 14429.7$ (Δ y ξ remain unalterable because they do not depend on the parameter a). From Theorem 1, it follows that $T^* = T_a = 4.23084$, $\Psi^* = a = 1.5$ and from Corollary 2, $G(T^*, \Psi^*) = 40.8645$.

Example 3 We use the same data given in Example 1, but let us change the values of D and K to $D = 5$ and $K = 1100$. We obtain $\Delta = -5464$, $L_1(a) = -5446$ and $\theta = -4690$. Since $\Psi_2 = 9.87715$, we have $C_2 = 64.3857 < C_o = 74.1620$. Therefore, the optimal policy is given by (T_2, Ψ_2), where $T_2 = 35.6314$ and $\Psi_2 = 9.87715$, with $G(T^*, \Psi^*) = 10.6143$.

Example 4 We consider the same input data and parameters of Example 1, but change the value of D to $D = 100$. Now $\Delta = 9400$ y $\theta = 319000$. Applying Theorem 1, we see that the optimal policy is $(T_o, 0)$, where $T_o = 1.41421$, with profit $G(T^*, \Psi^*) = 4.28932$.

Example 5 We now assume the same input parameters as in Example 3, but modify the value of π to $\pi = 0$. We obtain $\xi = 0$ and $\theta = -5185$ (Δ remains unalterable because it does not depend on the parameter π). Applying again Theorem 1, we have $T^* = \Psi^* = \infty$ and, therefore, $G(T^*, \Psi^*) = 0$. Obviously, in this case, the item should not be stocked.

§6. Particular models

Next, we prove that the inventory system studied here generalizes several inventory models analyzed by other authors. So, we have the following results:

1. If we consider $a \to \infty$ then, at the limit, we have $\beta(\tau) = \rho$ for $\tau > 0$, and we revert to the model with fixed partial backlogging studied by San-José, Sicilia & Garcia-Laguna (2009). Now, we get $\lim_{a\to\infty} \theta = \infty$ and $\lim_{a\to\infty} L(a) = \begin{cases} \infty & \text{if } \xi > 0 \\ \Delta & \text{if } \xi = 0 \end{cases}$.

   Thus, applying the results shown in Table 1, our optimal policy coincides with the one obtained by those authors. Moreover, several particular cases are deduced:

   (a) If $\pi = 0$ and $\rho\omega > 0$, we obtain the model studied by Montgomery, Bazaraa & Keswani (1973) and Rosenberg (1979).

   (b) If $\pi = 0$ and $\omega_o = 0$, the problem is equivalent to the one studied by Park (1982), Chu & Chung (2004), Yang (2007) and Leung (2008).

2. If we suppose $\pi = 0$ and $\rho = 1$, then we revert to the model developed by San-Jose, Sicilia & Garcia-Laguna (2005).

§7. Conclusions

In this work we study a continuous review inventory control system with deterministic demand where the demand on the shortage period is partially backlogged. We consider that the customers are impatient and, therefore, a part of the demand during the stockout period may not be satisfied. We suppose that the fraction of the customers who are willing to wait for the next replenishment is a step function on the waiting time. We assume that the stockout costs (backorder cost and lost sales cost) are both made up of a fixed cost and a variable cost which linearly depends on the length of the waiting time for the next replenishment. We develop a procedure to determine the optimal policy and the maximum inventory profit. This model generalizes several known inventory systems.

Further extensions to the developed model can be done for demand rate depending on inventory level, deteriorating items, demand rate depending on selling price, production finite rate (EPQ model), etc. Also, we could generalize the inventory system considering that the cumulative stockout unit cost is described by an increasing and quadratic function of the waiting time up to the next replenishment. These extensive problems will be considered in our future research.

Acknowledgements

This work is partially supported by the Spanish Ministry of Education and Science, Research Projects MTM2007-60928 and MTM2010-18591, from the National Plan of Scientific Research, Technological Development and Innovation. Also, this work is partially supported by the project OTP/09/08 of the Regional Government of Castilla and León Community (Spain).

References

[1] CHERN, M.S., CHAN, Y.L., AND TENG, J.T. A comparison among various inventory shortage models for deteriorating items on the basic of maximising profit. *Asia-Pacific Journal of Operational Research 22*, 1 (2005), 121–134.

[2] CHU, P., AND CHUNG, K.J. The sensitivity of the inventory model with partial backorders. *European Journal of Operational Research 152*, 1 (2004), 289–295.

[3] LEE, H.L., AND NAHMIAS, S. Single-product single-location models. In *Handbooks in Operations Research and Management Science vol. 4*, Graves, S.C., Rinnoy, A.H.C., and Zipkin, P.H. (editors), North-Holland, Amsterdam, 1993, 3–55.

[4] LEUNG, K.N. Technical Note: A use of the complete squares method to solve and analyze a quadratic objective function with two decision variables exemplified via deterministic inventory model with a mixture of backorders and lost sales. *International Journal of Production Economics 113*, 1 (2008), 275–281.

[5] MONTGOMERY, D.C., BAZARAA, M.S., AND KESWANI, A.K. Inventory models with a mixture of backorders and lost sales. *Naval Research Logistics Quarterly 20*, 2 (1973), 255–263.

[6] PARK, K.S. Inventory model with partial backorders. *International Journal of Systems Science 13*, 12 (1982), 1313–1317.

[7] ROSENBERG, D. A new analysis of a lot size model with partial backlogging. *Naval Research Logistics Quarterly 26*, 2 (1979), 346–353.

[8] SAN JOSÉ, L.A., SICILIA, J., AND GARCÍA-LAGUNA, J. The lot size-reorder level inventory system with customers impatience functions. *Computers & Industrial Engineering 49*, 3 (2005), 349–362.

[9] SAN JOSÉ, L.A., SICILIA, J., AND GARCÍA-LAGUNA, J. A general model for EOQ inventory systems with partial backlogging and linear shortage costs. *International Journal of Systems Science 40*, 1 (2009), 59–71.

[10] YANG, G.K. Note on sensitivity analysis of inventory model with partial backorders. *European Journal of Operational Research 177*, 2 (2007), 865–871.

Luis-A. San-José
Departamento de Matemática Aplicada
E.T.S. de Ingeniería Informática
Universidad de Valladolid
augusto@mat.uva.es

Joaquín Sicilia
Departamento de Estadística, Investigación Operativa y Computación
Facultad de Matemáticas
Universidad de La Laguna
jsicilia@ull.es

Juan García-Laguna
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias
Universidad de Valladolid
laguna@eio.uva.es

Monografías Matemáticas García de Galdeano **36**, 147–156 (2010)

ON CLASSIFICATION OF MINIMA AND MAXIMA FROM SOME BIVARIATE DISTRIBUTIONS WITH BOUNDED SUPPORT

Juana-María Vivo and Manuel Franco

Abstract. Some well-known continuous models with bounded support have been studied in the literature, such as uniform and triangular distributions. Despite its long history, the recognition of its importance as analytical tools has been relatively recent (see e.g. Kotz and van Dorp (2004)).

Unfortunately, a univariate distribution does not have a unique extension to the bivariate case. Several bivariate models have been proposed for a univariate model (see e.g. Kotz et al. (2000)), the most of them defined on either the whole $\mathbb{R}^2$ or on a positive orthant. Recently, a bivariate triangular model under independence have been used in Glickman and Xu (2008).

Furthermore, the log-concavity properties of the distribution models have interesting qualitative implications in many fields, a good review and applications can be found in Bagnoli and Bergstrom (2005). Therefore, the main aim of this paper is focused on the log-concavity properties associated to some continuous bivariate distributions on a bounded support, which can be considered as bivariate extensions of models on a bounded interval.

Keywords: Bounded models, Bivariate distributions, Log-concavity

AMS classification: 62H10, 62N05, 91B02

§1. Introduction

The minimum and maximum order statistics play an important role in various statistical applications in which most of the systems have dependent structures. For instance, in survival analysis with different unobserved causes, the minimum is the observable time of death in a competing risks model, and the maximum represents the observable time of death in a complementary risks model. In reliability theory, minima and maxima can be viewed as the lifetimes of series and parallel systems, which are determined by the working of all or at least one of the components, respectively. Thus, their ageing classifications or log-concavity properties are of interest in that scheduling.

In general, the log-concavity of the survival function is defined by the conditional survival function of a unit of age x: $S(t|x) = S(t+x)/S(x)$ where $S(x) = P(X > x)$ is the survival or reliability function of a random variable (rv) X, i.e., the survival probability over

an additional period of duration t of a unit of age x. Here, the physical principle of no ageing associated to the exponential, also called loss of memory property (LMP) $S(t|x) = S(t)$, represents that it does not age stochastically, i.e. its survival probability over an additional period of duration t is the same regardless of its present age x. So, the monotonicity of the conditional survival function with respect to its actual age points out the ageing of the unit, and then it is said that a nonnegative rv X is increasing (decreasing) failure rate, IFR (DFR), if $S(t|x)$ is decreasing (increasing) in x for each $t \geq 0$. A good reference for these ageing notions is the text of Barlow and Proschan (1981).

Note that the log-concavity of the survival can be analyzed through the density function (pdf) since the log-concavity of the density function implies the log-concavity of the survival function (see Ross (1996) and Franco and Vivo (2002) among others). Specially, the log-concave density function class includes among others the normal, truncated normal, exponential, Weibull, gamma, beta, uniform and Laplace. In some of these models, the log-concavity requires restrictions on the parameter values. For example, the proofs and constraints for exponential, gamma and Laplace distributions are derived from their definitions, see Johnson et al. (1994). Likewise, the parametric restrictions to ensure the log-concavity of generalized mixtures have been studied in Baggs and Nagaraja (1996) and Franco and Vivo (2002, 2006, 2007, 2009). A list of distributions with log-concave density functions in its support can be found in Bagnoli and Bergstrom (2005) wherein only the uniform and beta distributions have bounded domain. Also, Bagnoli and Bergstrom (2005) provide an excellent review on the log-concavity properties and several applications with interesting qualitative implications in many areas of economics, political science, actuarial science, biology and engineering, in which explicit assumptions on the underlying distribution are usually required.

Furthermore, bivariate data frequently arise in real life applications, and in these situations it is important to consider different bivariate distributions. Most of the bivariate distributions have been constructed as extensions of univariate distributions; unfortunately, these extensions are not unique (see e.g. Kotz et al. (2000) and Balakrishnan and Lai (2009)). Moreover, many of these applications are associated to a bounded domain, which are usually analyzed by using unbounded distributions, although bounded models can be more appropriate in some situations (see e.g. Kotz and van Dorp (2004)). Thus, it seems reasonable to use bivariate distributions with bounded support as underlying distributions in those problems. For example, in reliability engineering, an association measure is used between an external factor and the failure of each unit of a two-component system, such as the ratio of "how much the external factor increases the probability of failure" compared with "how much an always fatal factor would increase the probability of failure". The joint support of both ratios is the unit square. Therefore, the main aim of this work is focused on the log-concavity properties of minima and maxima from some bivariate distributions with bounded support. Note that the log-concavity of the minimum order statistic from a bivariate model is related with the weak bivariate ageing notions.

The paper is organized as follows. Section 2 displays some definitions and properties of the log-concavity or ageing classes. Section 3 introduces some usual continuous bivariate distributions with bounded support which can be considered as bivariate extensions of univariate models. The extreme order statistics from these bivariate models are given, and then, their log-concavities are also classified in this section.

§2. Preliminaries

In the first place, we show some concepts and previous results, which will be used in the remaining sections.

Let (X_1, X_2) be a generic two-dimensional random vector, whose components are neither necessarily independent nor identically distributed, with cumulative distribution functions $F_i(x)$, survival functions $S_i(x)$, and pdf's $f_i(x)$, $i = 1, 2$. Let $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$ be its joint distribution function, and $S(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$ its survival function. Some continuous bivariate models can be seen in Kotz et al. (2000) and Balakrishnan and Lai (2009).

Let $T_1 = \min(X_1, X_2)$ and $T_2 = \max(X_1, X_2)$ be the minimum and maximum order statistics, and $S_{(i)}(x)$ and $f_{(i)}(x)$ denote their survival and density functions, $i = 1, 2$, respectively. Thus, the survival functions of both extreme statistics are given by

$$S_{(1)}(x) = S(x, x) = P(X_1 > x, X_2 > x) \tag{1}$$

and

$$S_{(2)}(x) = S_1(x) + S_2(x) - S_{(1)}(x). \tag{2}$$

On the other hand, let X be a non-negative rv with survival function $S(x)$. It is said that X has a log-concave (log-convex) survival function, if $\log S(x)$ is concave (convex) in its support. In the absolutely continuous case, the log-concavity of the survival function is determined by the monotonicity of its failure rate function, wherein the failure rate function $r(x) = -\frac{d}{dx} \log S(x) = \frac{f(x)}{S(x)}$ represents the probability of failure or death in each moment. So, the log-concave (log-convex) survival function is also well-known by IFR (DFR) ageing class. Now, let us see a technical lemma which is often used to study the log-concavity of the survival function, see e.g. Barlow and Proschan (1981).

Lemma 1. *Let X be an absolutely continuous rv with survival function $S(x)$ and density function $f(x)$. If*

$$f'(x) \cdot S(x) + f(x)^2 \geq (\leq) 0$$

for all x, then X has a log-concave (log-convex) survival function, i.e., X is IFR (DFR).

Likewise, it is said that X has a log-concave (log-convex) density function, if $\log f(x)$ is concave (convex) in its support. It is well-known by increasing (decreasing) likelihood ratio, shortly denoted by ILR (DLR) ageing class, wherein the likelihood ratio function $l(x) = -\frac{d}{dx} \log f(x)$. Thus, the log-concavity of the pdf is determined by the monotonicity of the likelihood ratio function, see e.g. Ross (1996). Note that DLR class only makes sense for a random variable with support not upper bounded. In practice the intuitive meaning of the assumption that a pdf is log-concave is that: (a) it does not have multiple separate maxima (although it could be flat on top), and (b) the tails of the density function are not "too thick". The following technical lemma will be used to explore the log-concavity and log-convexity of the pdf, see e.g. Franco and Vivo (2002).

Lemma 2. *Let X be an absolutely continuous rv with density function $f(x)$. If*

$$f''(x) \cdot f(x) \leq (\geq) f'(x)^2$$

for all x, then X has log-concave (log-convex) pdf, i.e., X is ILR (DLR).

Moreover, the log-concavity of the survival and density functions satisfies the following relationships, see Ross (1996) and Franco and Vivo (2002) among others.

Lemma 3. *Let X be an absolutely continuous rv. If the pdf is log-concave (log-convex) in its support then the survival function is log-concave (log-convex) in its support.*

Finally, we shall use the following technical lemma about the concavity (convexity) of the piecewise differentiable function in its support, see e.g. Vivo and Franco (2008).

Lemma 4. *Let $g(x)$ be a real continuous and piecewise differentiable function in its support. If $g(x)$ is piecewise concave and $g'(x-) \geq g'(x+)$, then it is concave in its support. Analogously, if $g(x)$ is piecewise convex and $g'(x-) \leq g'(x+)$, then it is convex in its support.*

§3. Minima and maxima from bivariate models with bounded support

In this section, we analyze three bivariate probability models on a bounded domain which have been utilized for the treatment of the uncertainty in valuation methodology (see Herrerías (2006)). They can be considered as bivariate extensions of continuous univariate distributions with bounded support which arise in several papers dealing with the PERT and risk analysis. Without loss of generality, we will consider their standardized versions, i.e., on the unit square $(0,1) \times (0,1) \subset \mathbb{R}^2$.
The minimum and maximum order statistics from these models are obtained. The log-concavity of their survival functions is also established, and then for their pdf's.

3.1. Cubic model

Definition 1. Let (X_1, X_2) be a bivariate rv. It is said that (X_1, X_2) follows a cubic model on $(0,1) \times (0,1)$, if its joint pdf is given by

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The cubic model or bivariate rectangular distribution on $(0,1) \times (0,1)$ is also known as bivariate uniform model, which is formed by two independent and identically distributed uniform components, and its joint survival function is easily obtained.
Besides, from (1) and (2), the survival functions of the minimum and maximum statistics from a cubic model can be expressed as

$$S_{(1)}(x) = \begin{cases} 1 & \text{if } x < 0 \\ (1-x)^2 & \text{if } 0 \leq x < 1 \\ 0 & \text{if } 1 \leq x \end{cases} \quad \text{and} \quad S_{(2)}(x) = \begin{cases} 1 & \text{if } x < 0 \\ 1 - x^2 & \text{if } 0 \leq x < 1 \\ 0 & \text{if } 1 \leq x. \end{cases} \tag{3}$$

Proposition 5. *The minimum and maximum order statistics from a cubic model have log-concave survival functions in its support.*

Proof. The proof follows from Lemma 1 and (3), taking into account that

$$f_{(1)}(x) = \begin{cases} 2(1-x) & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad f_{(2)}(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

are the pdf's of the minimum and maximum statistics, respectively. □

Proposition 6. *The minimum and maximum order statistics from a cubic model have log-concave pdf's in its support.*

Proof. It is obvious from Lemma 2 and the pdf's given in the proof of Proposition 5. □

3.2. Rectangular-triangular model

Definition 2. Let (X_1, X_2) be a bivariate rv. It is said that (X_1, X_2) has a rectangular-triangular model on $(0,1) \times (0,1)$, with parameter $m \in (0,1)$, if its joint pdf is given by

$$f(x_1, x_2) = \begin{cases} 2\frac{x_2}{m} & \text{if } 0 < x_1 < 1,\ 0 < x_2 < m \\ 2\frac{1-x_2}{1-m} & \text{if } 0 < x_1 < 1,\ m \leq x_2 < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The rectangular-triangular model corresponds to a bivariate rv with independent components. One of them is uniformly distributed and another component has a triangular distribution with parameter m, and hence, its joint survival function is easily obtained.
Moreover, from (1) and (2), the survival functions of their extreme statistics can be written as

$$S_{(1)}(x) = \begin{cases} 1 & \text{if } x < 0 \\ \frac{1}{m}(1-x)(m-x^2) & \text{if } 0 \leq x < m \\ \frac{1}{1-m}(1-x)^3 & \text{if } m \leq x < 1 \\ 0 & \text{if } 1 \leq x \end{cases} \tag{4}$$

and

$$S_{(2)}(x) = \begin{cases} 1 & \text{if } x < 0 \\ 1 - \frac{1}{m}x^3 & \text{if } 0 \leq x < m \\ 1 - x\left(1 - \frac{1}{1-m}(1-x)^2\right) & \text{if } m \leq x < 1 \\ 0 & \text{if } 1 \leq x. \end{cases} \tag{5}$$

Proposition 7. *The minimum and maximum order statistics from a rectangular-triangular model have log-concave survival functions in its support.*

Proof. From (4) and (5), the pdf's of the minimum and maximum statistics from a rectangular-triangular model are given by

$$f_{(1)}(x) = \begin{cases} 1 - \frac{x}{m}(3x-2) & \text{if } 0 < x < m \\ \frac{3}{1-m}(1-x)^2 & \text{if } m \leq x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$f_{(2)}(x) = \begin{cases} \frac{3}{m}x^2 & \text{if } 0 < x < m \\ 1 - \frac{1}{1-m}(1-x)(1-3x) & \text{if } m \leq x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and from Lemma 1, $S_{(1)}(x)$ and $S_{(2)}(x)$ are both log-concave in each interval, i.e., $\log S_{(1)}(x)$ and $\log S_{(2)}(x)$ are piecewise concave in $(0,1)$. Moreover, their failure rate functions can be expressed as

$$r_{(1)}(x) = -\frac{d}{dx}\log S_{(1)}(x) = \begin{cases} \frac{1}{1-x} + \frac{2x}{m-x^2} & \text{if } 0 < x < m \\ \frac{3}{1-x} & \text{if } m \leq x < 1 \end{cases}$$

and

$$r_{(2)}(x) = -\frac{d}{dx}\log S_{(2)}(x) = \begin{cases} \frac{3x^2}{m-x^3} & \text{if } 0 < x < m \\ \frac{1}{1-x} + \frac{2x-1}{1-m+x(1-x)} & \text{if } m \leq x < 1. \end{cases}$$

Thus, we have that $r_{(1)}(m-) = r_{(1)}(m+)$ and $r_{(2)}(m-) = r_{(2)}(m+)$, and consequently, from Lemma 4, both survival functions are log-concave in $(0,1)$. □

Proposition 8. *The minimum and maximum order statistics from a rectangular-triangular model have log-concave pdf's in its support.*

Proof. From the pdf's given in the proof of Proposition 7, it is easy to prove that both pdf's are piecewise log-concave in $(0,1)$ by Lemma 2. Moreover, we have that

$$\big(\log f_{(1)}(x)\big)' = \frac{f'_{(1)}(x)}{f_{(1)}(x)} = \begin{cases} -2\frac{3x-1}{m-x(3x-2)} & \text{if } 0 < x < m \\ -\frac{2}{1-x} & \text{if } m \leq x < 1 \end{cases}$$

and

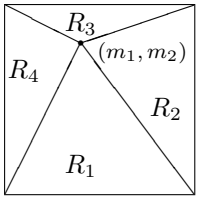
$$\big(\log f_{(2)}(x)\big)' = \frac{f'_{(2)}(x)}{f_{(2)}(x)} = \begin{cases} \frac{2}{x} & \text{if } 0 < x < m \\ \frac{2(2-3x)}{1-m-(1-x)(1-3x)} & \text{if } m \leq x < 1 \end{cases}$$

and from Lemma 4, we obtain that the pdf's of the minimum and maximum statistics are log-concave in $(0,1)$, since $(\log f_{(i)}(m-))' \geq (\log f_{(i)}(m+))'$, $i = 1, 2$. □

3.3. Pyramidal model

Definition 3. Let (X_1, X_2) be a bivariate rv. It is said that (X_1, X_2) follows a pyramidal model on $(0,1)\times(0,1)$, with parameters (m_1, m_2), such that $m_1, m_2 \in (0,1)$, if its joint pdf is given by

$$f(x_1,x_2) = \begin{cases} \frac{3}{m_2}x_2 & \text{if } (x_1,x_2) \in R_1 \\ \frac{3}{1-m_1}(1-x_1) & \text{if } (x_1,x_2) \in R_2 \\ \frac{3}{1-m_2}(1-x_2) & \text{if } (x_1,x_2) \in R_3 \\ \frac{3}{m_1}x_1 & \text{if } (x_1,x_2) \in R_4 \\ 0 & \text{elsewhere} \end{cases}$$



where the regions R_i's are defined by the following triangles of the unit square

$$
\begin{aligned}
R_1 &= \left\{(x_1,x_2):\ \tfrac{m_1}{m_2}x_2 \le x_1 \le 1-\tfrac{(1-m_1)}{m_2}x_2,\ 0<x_2\le m_2\right\}\\
R_2 &= \left\{(x_1,x_2):\ m_1 \le x_1 < 1,\ \tfrac{m_2(1-x_1)}{1-m_1} \le x_2 \le 1-\tfrac{(1-m_2)(1-x_1)}{1-m_1}\right\}\\
R_3 &= \left\{(x_1,x_2):\ \tfrac{m_1(1-x_2)}{1-m_2} \le x_1 \le 1-\tfrac{(1-m_1)(1-x_2)}{1-m_2},\ m_2\le x_2<1\right\}\\
R_4 &= \left\{(x_1,x_2):\ 0<x_1\le m_1,\ \tfrac{m_2}{m_1}x_1 \le x_2 \le 1-\tfrac{1-m_2}{m_1}x_1\right\}.
\end{aligned}
$$

The joint survival function of a pyramidal model can be obtained by $S(x_1,x_2)=F(x_1,x_2)-F(x_1,1)-F(1,x_2)+1$, being $F(x_1,x_2)$ its joint distribution function

$$
F(x_1,x_2)=\begin{cases}
0 & \text{if } x_1\le 0 \text{ or } x_2\le 0\\
\frac{3x_1x_2^2}{2m_2}-\frac{m_1x_2^3}{2m_2^2} & \text{if } (x_1,x_2)\in R_1\\
\frac{m_2(1-x_1)^3}{2(1-m_1)^2}-\frac{x_2^3}{2m_2^2}+\frac{3x_2^2}{2m_2}-\frac{3x_2(1-x_1)^2}{2(1-m_1)} & \text{if } (x_1,x_2)\in R_2,\ x_2<m_2\\
1+\frac{(1-x_2)^3}{2(1-m_2)^2}+\frac{m_2(1-x_1)^3}{2(1-m_1)^2}-\frac{3(1-x_2)^2}{2(1-m_2)} & \\
\quad -\frac{3x_2(1-x_1)^2}{2(1-m_1)} & \text{if } (x_1,x_2)\in R_2,\ m_2\le x_2<1\\
\frac{m_1(1-x_2)^3}{2(1-m_2)^2}-\frac{x_1^3}{2m_1^2}+\frac{3x_1^2}{2m_1}-\frac{3x_1(1-x_2)^2}{2(1-m_2)} & \text{if } (x_1,x_2)\in R_3,\ x_1<m_1\\
1+\frac{(1-x_1)^3}{2(1-m_1)^2}+\frac{m_1(1-x_2)^3}{2(1-m_2)^2}-\frac{3(1-x_1)^2}{2(1-m_1)} & \\
\quad -\frac{3x_1(1-x_2)^2}{2(1-m_2)} & \text{if } (x_1,x_2)\in R_3,\ m_1\le x_1<1\\
\frac{3x_1^2x_2}{2m_1}-\frac{m_2x_1^3}{2m_1^2} & \text{if } (x_1,x_2)\in R_4\\
\frac{3x_1^2}{2m_1}-\frac{x_1^3}{2m_1^2} & \text{if } 0\le x_1<m_1,\ 1\le x_2\\
1+\frac{(1-x_1)^3}{2(1-m_1)^2}-\frac{3(1-x_1)^2}{2(1-m_1)} & \text{if } m_1\le x_1<1,\ 1\le x_2\\
\frac{3x_2^2}{2m_2}-\frac{x_2^3}{2m_2^2} & \text{if } 0\le x_2<m_2,\ 1\le x_1\\
1+\frac{(1-x_2)^3}{2(1-m_2)^2}-\frac{3(1-x_2)^2}{2(1-m_2)} & \text{if } m_2\le x_2<1,\ 1\le x_1\\
1 & \text{if } 1\le x_1,\ 1\le x_2.
\end{cases}
$$

Note that the pyramidal model can be considered as a more appropriate bivariate version of the triangular model in dependent scenarios.

Furthermore, using the notation $m=\min(m_1,m_2)$ and $M=\max(m_1,m_2)$, from (1) and (2), and taking into account that the diagonal of unit square intersects the line between the triangles R_1 and R_2 in $x=\frac{m_2}{1+m_2-m_1}$, and the line between the triangles R_3 y R_4 in $x=\frac{m_1}{1+m_1-m_2}$, the survival functions of the minimum and maximum statistics from a

pyramidal model can be expressed as

$$S_{(1)}(x)=\begin{cases}1 & \text{if } x<0\\ 1-\frac{x^2(3m-x)}{2m^2}-\frac{x^2(3M-x)}{2M^2}+\frac{x^3(3M-m)}{2M^2} & \text{if } 0\le x<m\\ \frac{(2+x-3m)(1-x)^2}{2(1-m)^2}+\frac{x^3(3M-m)}{2M^2}-\frac{x^2(3M-x)}{2M^2} & \text{if } m\le x<\frac{M}{1+M-m}\\ \frac{(2+M-3m)(1-x)^3}{2(1-m)^2} & \text{if } \frac{M}{1+M-m}\le x<1\\ 0 & \text{if } 1\le x\end{cases} \tag{6}$$

and

$$S_{(2)}(x)=\begin{cases}1 & \text{if } x<0\\ 1-\frac{x^3(3M-m)}{2M^2} & \text{if } 0\le x<\frac{M}{1+M-m}\\ 1-\frac{M(1-x)^3}{2(1-m)^2}-\frac{3x^2}{2M}+\frac{x^3}{2M^2}+\frac{3x(1-x)^2}{2(1-m)} & \text{if } \frac{M}{1+M-m}\le x<M\\ \frac{3(1-x)^2}{2(1-M)}+\frac{3x(1-x)^2}{2(1-m)}-\frac{(1-x)^3}{2(1-M)^2}-\frac{M(1-x)^3}{2(1-m)^2} & \text{if } M\le x<1\\ 0 & \text{if } 1\le x.\end{cases} \tag{7}$$

Now, we obtain the log-concavity of the pdf's of the minimum and maximum order statistics from a pyramidal model, and consequently, from Lemma 3 we shall have the log-concavity of the survival functions of minima and maxima from this model.

Proposition 9. *The minimum and maximum order statistics from a pyramidal model have log-concave pdf's in its support.*

Proof. From the survival functions (6) and (7) of the minimum and maximum statistics from a pyramidal model, their corresponding pdf's can be written as

$$f_{(1)}(x)=\begin{cases}\frac{3x\left(2mM(M+m)-(M^2+m^2(3M-m+1))x\right)}{2m^2M^2} & \text{if } 0<x<m\\ \frac{3(m-3M-1)x^2}{2M^2}-\frac{3Mx^2-6(mM+(1-m)^2)x-3(1-2m)M}{2(1-m)^2M} & \text{if } m\le x<\frac{M}{1+M-m}\\ \frac{3(2-3m+M)(1-x)^2}{2(1-m)^2} & \text{if } \frac{M}{1+M-m}\le x<1\\ 0 & \text{elsewhere}\end{cases}$$

and

$$f_{(2)}(x)=\begin{cases}\frac{3x^2(3M-m)}{2M^2} & \text{if } 0<x\le\frac{M}{1+M-m}\\ \frac{3x}{M}-\frac{3M(1-x)^2}{2(1-m)^2}-\frac{3x^2}{2M^2}-\frac{3(1-4x+3x^2)}{2(1-m)} & \text{if } \frac{M}{1+M-m}<x\le M\\ \frac{3(1-x)}{1-M}-\frac{3(1-x)^2}{2(1-M)^2}-\frac{3M(1-x)^2}{2(1-m)^2}-\frac{3(1-4x+3x^2)}{2(1-m)} & \text{if } M<x\le 1\\ 0 & \text{elsewhere.}\end{cases}$$

Taking into account the first and second derivatives of both pdf's in each interval, from Lemma 2, the log-concavity of $f_{(i)}(x)$ $(i=1,2)$ is determined by the signs of the numerators

of the first derivative of their likelihood ratio functions

$$l'_{(1)}(x) = \begin{cases} \frac{F^2x^2+(Fx-2mM)^2}{(2mM(M+m)x-Fx^2)^2} & \text{if } 0 < x < m \\ \frac{2H(1-2m)M^2+x^2H^2+(G-xH)^2}{((1-2m)M^2+2Gx-Hx^2)^2} & \text{if } m \leq x < \frac{M}{1+M-m} \\ \frac{2}{(1-x)^2} & \text{if } \frac{M}{1+M-m} \leq x < 1 \end{cases}$$

and

$$l'_{(2)}(x) = \begin{cases} \frac{2}{x^2} & \text{if } 0 < x < \frac{M}{1+M-m} \\ \frac{2\left(Ax+M(M+1-m)^2\right)^2+2M^2(M+1-m)(1-m)^2(3M-m)}{\left(Ax^2-2M(M+1-m)^2x+M^2(M+1-m)\right)^2} & \text{if } \frac{M}{1+M-m} \leq x < M \\ \frac{2\left((Cx-B)^2+B^2+2CD\right)}{(2D+2Bx-Cx^2)^2} & \text{if } M \leq x < 1 \end{cases}$$

where

$$\begin{aligned} A &= M^3 + (1-m)^2 + 3(1-m)M^2 \\ B &= M(M-m)^2 + 2(1-M)(1-m) \\ C &= (1-m)^2 + (1-M)^2(M+3(1-m)) \\ D &= \tfrac{1}{2}\left((1-m)^2(1-2M) + (1-M)^2(1-M-m)\right) \\ F &= M^2 + m^2(3M-m+1) \\ G &= M\left(mM + (1-m)^2\right) \\ H &= M^2 + (1-m)^2(3M-m). \end{aligned}$$

Therefore, $f_{(i)}(x)$ is log-concave in each interval of its support, $i = 1, 2$. Moreover, $\log f_{(i)}(x)$ verifies the conditions of Lemma 4, $i = 1, 2$, since $l'_{(1)}(m-) = l'_{(1)}(m+)$, $l'_{(1)}(\frac{M}{1+M-m}-) = l'_{(1)}(\frac{M}{1+M-m}+)$, $l'_{(2)}(\frac{M}{1+M-m}-) = l'_{(2)}(\frac{M}{1+M-m}+)$ and $l'_{(2)}(M-) = l'_{(2)}(M+)$. So, we have that the pdf's of the minimum and maximum statistics are log-concave in $(0, 1)$. □

Corollary 10. *The minimum and maximum order statistics from a pyramidal model have log-concave survival functions in its support.*

Proof. It is obvious from Lemma 3 and Proposition 9. □

Acknowledgements

The authors would like to thank the referee for the comments and careful reading of the paper. This work was partially supported by Fundación Séneca (CARM) under grant 11886/PHCS/09.

References

[1] BAGGS, G.E., AND NAGARAJA, H.N. Reliability properties of order statistics from bivariate exponential distributions. *Comm. Statist. Stoch. Models 12* (1996), 611–631.

[2] BAGNOLI, M., AND BERGSTROM, T. Log-concave probability and its applications. *Economic Theory 26* (2005), 445–469.

[3] BALAKRISHNAN, N., AND LAI, C.D. *Continuous Bivariate Distributions (2nd ed.)*. Springer, New York, 2009.

[4] BARLOW, R.E., AND PROSCHAN, F. *Statistical Theory of Reliability and Life Testing: Probability Models*. MD: McArdale Press, Silver Springs, Maryland, 1981.

[5] FRANCO, M., AND VIVO, J.M. Reliability properties of series and parallel systems from bivariate exponential models. *Comm. Statist. Theory Meth. 31* (2002), 2349–2360.

[6] FRANCO, M., AND VIVO, J.M. On Log-concavity of the extremes from Gumbel bivariate exponential distributions. *Statistics 40* (2006), 415–433.

[7] FRANCO, M., AND VIVO, J.M. Generalized mixtures of gamma and exponentials and reliability properties of the maximum from Friday and Patil bivariate exponential model. *Comm. Statist. Theory Meth. 36* (2007), 2011–2025.

[8] FRANCO, M., AND VIVO, J.M. Stochastic aging classes for the maximum statistic from Friday and Patil bivariate exponential distribution family. *Comm. Statist. Theory Meth. 38* (2009), 902–911.

[9] GLICKMAN, T.S., AND XU, F. The distribution of the product of two triangular random variables. *Statistic and Probability Letters 78* (2008), 2821–2826.

[10] HERRERÍAS, J.M. *Modelos probabilísticos aplicados a la Teoría de Valoración: El método de las dos betas*. Fundación Unicaja, Málaga, 2006.

[11] JOHNSON, N.L., KOTZ, S., AND BALAKRISHNAN, N. *Continuous Univariate Distributions (Vol. 1, 2nd ed.)*. Wiley, New York, 1994.

[12] KOTZ, S., BALAKRISHNAN, N., AND JOHNSON, N.L. *Continuous Multivariate Distributions*. Wiley, New York, 2000.

[13] KOTZ, S., AND VAN DORP, J.R. *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*. World Scientific Press, Singapore, 2004.

[14] ROSS, S. *Stochastic Processes (2nd ed.)*. Wiley, New York, 1996.

[15] VIVO, J.M., AND FRANCO, M. On stochastic classification of the generalized trapezoidal distribution family. In *Monografías del Seminario Matemático García de Galdeano 34*, B. Lacruz et al. Eds. Prensas Universitarias de Zaragoza, 2008, 197–204.

Juana-María Vivo
Dept. Quantitative Methods for Economy
University of Murcia
30100 Murcia. Spain
jmvivomo@um.es

Manuel Franco
Dept. Statistics and Operations Research
University of Murcia
30100 Murcia. Spain
mfranco@um.es

Authors

OTHER COMMUNICATIONS

The following contributions are the ones which were presented but not included in this book. Some will appear in other publications.

Trend analysis of water quality series based on regression models with correlated errors.
Jesús Abaurrea, Jesús Asín, Ana C. Cebrián and Miguel A. García-Vera

A new heuristic to solve the one-warehouse N-retailer problem.
Beatriz Abdul-Jalbar, Anders Segerstedt, Joaquín Sicilia and Andreas Nilsson

Solving a stochastic sequential ordering problem for minimizing the expected number of tardy jobs.
David Alcaide López de Pablo, Andrés Rodríguez González and Joaquín Sicilia Rodríguez

Dependent Data (Re)Combination: An order statistics approach
Adolfo Alvarez Pinto and Daniel Peña Sánchez de Rivera

Evaluating the predictiveness of new biomarkers in prostate cancer recurrence
David Arteta, Luis M. Esteban, Jokin del Amo, Diego Tejedor, Angel Borque and Gerardo Sanz

Statistical analysis of input processes and length of stay in intensive care units
Cristina Azcárate, Fermín Mallor and Susana Ochoa

Exact Prediction Intervals for Future Exponential Lifetime Based on Random Generalized Order Statistics
H. M. Barakat, Magdy E. El-Adll and Amany E. Aly

On Multivariate Order Statistics. Asymptotic Theory and Computing Moments
H.M. Barakat and E.M. Nigm

Bootstrapping extreme generalized order statistics
H.M. Barakat, E.M. Nigm and Magdy E. El-Adll

Robustness measures for estimators in non-regular distributions
Inmaculada Barranco-Chamorro and Juan Luis Moreno-Rebollo

Bayesian network classifiers: An application to predict the success of a smoking cessation therapy
María Asunción Beamonte, Pilar Gargallo, Adriana Jiménez-Muro, Beatriz Lacruz, Adriana Marqueta and Isabel Nerín

On random contractions of generalized order statistics
Milena Bieniek

Some new complexity results on fuzzy integer programming
Víctor Blanco and Justo Puerto

On the optimization of the quotient of DC functions
Rafael Blanquero and Emilio Carrizosa

Robust estimation of the scale parameter in weighted models
Paweł Błażej

A degradation model based on a gamma process and Brownian motion
Laurent Bordes, Christian Paroissin and Ali Salami

Convolution order for records and non-linear Volterra integral equations
Antonia Castaño-Martínez, José Fernando López-Blázquez and Begoña Salamanca-Miño

Order polytopes versus subfamilies of fuzzy measures
Elías F. Combarro and Pedro Miranda

Studying for operating problems of Mould and Die Industry in Thailand
Mongkon Deeudom and Surapun Yimman

The Efficiency of Population Mean Estimation and Population Proportion Estimation for Normal and Skewed Data
Boonkong Dhakonlayodhin

On the Change Point Estimation of Exponentially Distributed Quality Characteristics
Eralp Dogu

Disintegration, Construction and Formulation of a Rostering Problem by Mathematical Programming
Zeynep Filiz Eren Dogu, Müjgan Sagir Ozdemir and C. Cengiz Celikoglu

Bayesian analysis of returns at industry and firms levels in european countries
José L. Gallizo, Pilar Gargallo, Ramón Saladrigues and Manuel Salvador

The Performance of Forecasting Models for Non-Linear Financial data
Pathom Glannamtip and Kanyapat Pussara

Maximum number of blocked terminals and number of collisions in slotted CSMA/CD channels
Antonio Gómez-Corral

Asymptotic results for the number of near-records and their statistical applications
Raúl Gouet, F. Javier López and Gerardo Sanz

Parameter estimation in linear and non linear models with general heteroscedastic structure
Ali Hadi, Beatriz Lacruz and Ana Pérez-Palomares

Effect of Unit Root Tests on Prediction Intervals for an Unknown Mean Gaussian AR(1) Process with Additive Outliers
Wararit Panichkitkosolkul and Sa-aat Niwitpong

A comparison study of sampling strategies in correlated populations
Ioulia Papageorgiou

Hitting time distribution of a critical threshold by some degradation models
Christian Paroissin and Ali Salami

Development of stratified randomized response technique of mixed model for sensitive question
Pasakorn Phanrod and Pairote Stirayakorn

Parameter Estimation Methods for Three-Parameter Generalized Pareto Distribution with Outliers
Wichitra Phonyiem, Prateep Subkongdee and Pathom Glannamtip

Non standard location problems and their cost allocation
Justo Puerto

Some statistical procedures for boundary estimation and image analysis
Alberto Rodríguez Casal

Different formulations for hub location problems with order
Antonio Manuel Rodríguez-Chía, Justo Puerto and Ana Belén Ramos

Studying the organization of semi-arid ecosystems: searching for plant associations and functional traits patterns
Hugo Saiz, Beatriz Lacruz and Concepción López Alados

Estimation of linear models under AUC optimization criterion, a step by step procedure to combine diagnostic tests
Gerardo Sanz, Luis Mariano Esteban and Angel Borque

Bias Reduction of the Process Capability Indices
Jeerapa Sappakitkamjorn and Sa-aat Niwitpong

Combining martingale and integral equations approaches for finding optimal designs of EWMA procedure
Saowanit Sukparungsee and Yupaporn Areepong

Development of Hair Dye from Herbal Extracts
Kongtun Janphuk Sumonthip and Suracherdkaiti Wichai

Improvement of Randomized Response Model by Twice Sampling Method
Wichai Suracherdkiati and Pasakorn Phanrod

ORGANIZERS

MAJOR SPONSORS

SPONSORS

MONOGRAFÍAS DEL SEMINARIO MATEMÁTICO "GARCÍA DE GALDEANO"

Desde 2001, el Seminario ha retomado la publicación de la serie *Monografías* en un formato nuevo y con un espíritu más ambicioso. El propósito es que en ella se publiquen tesis doctorales dirigidas o elaboradas por miembros del Seminario, actas de Congresos en cuya organización participe o colabore el Seminario y monografías en general. En todos los casos, se someten al sistema habitual de arbitraje anónimo.

Los manuscritos o propuestas de publicaciones en esta serie deben remitirse a alguno de los miembros del Comité editorial. Los trabajos pueden estar redactados en español, francés o inglés.

Las monografías son recensionadas en *Mathematical Reviews* y en *Zentralblatt MATH*.

Últimos volúmenes de la serie:

21. A. Elipe y L. Floría (eds.): *III Jornadas de Mecánica Celeste*, 2001, ii + 202 pp., ISBN: 84-95480-21-2.

22. S. Serrano Pastor: *Modelos analíticos para órbitas de satélites artificiales de tipo quasi-spot*, 2001, vi + 76 pp., ISBN: 84-95480-35-2.

23. M. V. Sebastián Guerrero: *Dinámica no lineal de registros electrofisiológicos*, 2001, viii + 251 pp., ISBN: 84-95480-43-3.

24. Pedro J. Miana: *Cálculo funcional fraccionario asociado al problema de Cauchy*, 2002, 171 pp., ISBN: 84-95480-57-3.

25. Miguel Romance del Río: *Problemas sobre Análisis Geométrico Convexo*, 2002, xvii + 214 pp., ISBN: 84-95480-76-X.

26. Renato Álvarez – Nodarse: *Polinomios hipergeométricos y q–polinomios*, 2003, vi +341 pp., ISBN: 84-7733-637-7.

27. M. Madaune – Tort, D. Trujillo, M. C. López de Silanes, M. Palacios, G. Sanz (eds.): *VII Jornadas Zaragoza – Pau de Matemática Aplicada y Estadística*, 2003, xxvi +523 pp., ISBN: 84-96214-04-4.

28. Sergio Serrano Pastor: *Teorías analíticas del movimiento de un satélite artificial alrededor de un planeta. Ordenación asintótica del potencial en el espacio fásico*, 2003, 164 pp., ISBN: 84-7733-667-9.

29. Pilar Bolea Catalán: *El proceso de algebrización de organizaciones matemáticas escolares*, 2003, 260 pp., ISBN: 84-7733-674-1.

30. Natalia Boal Sánchez: *Algoritmos de reducción de potencial para el modelo posinomial de programación geométrica*, 2003, 232 pp., ISBN: 84-7733-667-9.

31. M. C. López de Silanes, M. Palacios, G. Sanz, J. J. Torrens, M. Madaune – Tort, D. Trujillo (eds.): *VIII Journées Zaragoza – Pau de Mathématiques Appliquées et de Statistiques*, 2004, xxvi +578 pp., ISBN: 84-7733-720-9.

32. Carmen Godes Blanco: *Configuraciones de nodos en interpolación polinómica bivariada*, 2006, xii +163 pp., ISBN: 84-7733-841-9.

33. M. Madaune – Tort, D. Trujillo, M. C. López de Silanes, M. Palacios, G. Sanz, J.J. Torrens (eds.): *Ninth International Conference Zaragoza–Pau on Applied Mathematics and Statistics*, 2006, xxxii +440 pp., ISBN: 84-7733-871-X.

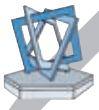
34. B. Lacruz, F.J. López, P. Mateo, C. Paroissin, A. Pérez-Palomares, y G. Sanz (eds.): *Pyrenees International Workshop on Statistics, Probability and Operations Research , SPO 2007*, 2008, 205 pp., ISBN: 978-84-92521-18-0.

35. M. C. López de Silanes, M. Palacios, G. Sanz, J. J. Torrens, M. Madaune – Tort, C. Paroissin, D. Trujillo (eds.): *Tenth International Conference Zaragoza–Pau on Applied Mathematics and Statistics*, 2010, xxx +272 pp., ISBN: 978-84-15031-53-6.

ISBN 978-84-15031925

Prensas Universitarias
Universidad Zaragoza

monografías
garcía de galdeano



Instituto Universitario de Investigación
de Matemáticas
y Aplicaciones
Universidad Zaragoza