MONOGRAFÍAS MATEMÁTICAS
"GARCÍA DE GALDEANO"

2013

N.º 38

The Pyrenees International Workshop and Summer School on Statistics, Probability and Operations Research SPO 2011

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 **38**, 2013

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

SPO 2011

Jaca, Spain, September 13-16, 2011

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 2011 : Jaca, Spain, September 13-16, 2011 / editors L. M. Esteban… [et al.]. — Zaragoza : Prensas de la Universidad de Zaragoza : Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, 2013
132 p. ; 24 cm. — (Monografías Matemáticas García de Galdeano ; 38)

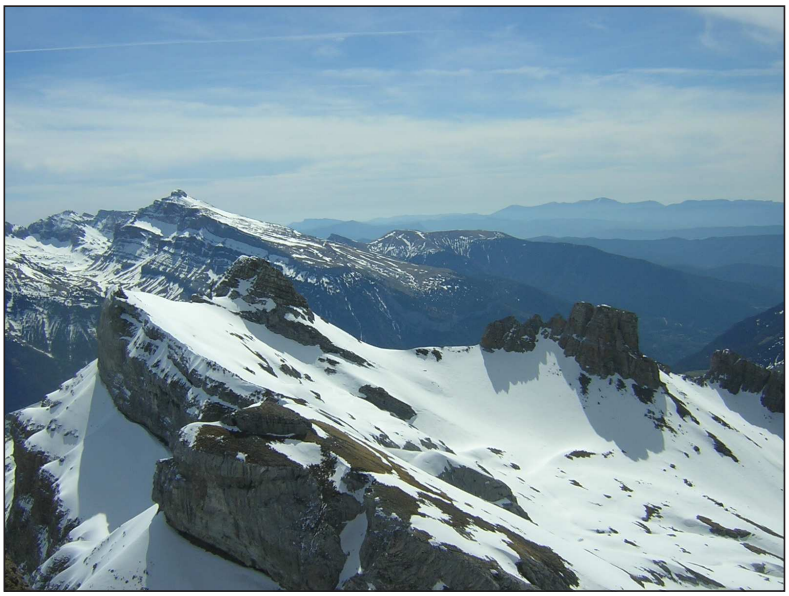
ISBN 978-84-15770-81-7
1. Estadística matemática. 2. Probabilidades
ESTEBAN, L. M.
519.21/.22

Monografías Matemáticas García de Galdeano n.º **38**
Octubre de 2013
Universidad de Zaragoza

Imprime: Servicio de Publicaciones. Universidad de Zaragoza

Depósito Legal: Z 1450-2013
ISBN: 978-84-15770-81-7

The edition of this volume has been partially subsidized by the Vicerrectorado de Investigación de la Universidad de Zaragoza, the grant MTM2011-15044-E of MICINN, the Government of Aragon, the project i-MATH and CTP.



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

Jaca (Huesca), Spain, 13-16 September, 2011

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).

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 2011, was held in Jaca (Spanish Pyrenees) from September 13 to September 16, 2011.

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

The school featured two advanced courses taught by Narayanaswamy Balakrishnan from the Department of Mathematics and Statistics, McMaster University, Canada (Precedence-type testing and applications) and Jesús López Fidalgo from the University of Castilla-La Mancha, Spain (Design of experiments for nonlinear models), and two plenary conferences by professor María Ángeles Gil from the University of Oviedo, Spain (Random fuzzy sets: a probabilistic tool to develop statistics with imprecise data) and professor Ali S. Hadi from The American University in Cairo, Egypt (Multi-class data exploration using space transformed visualization plots).

We thank them very sincerely.

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

This volume includes 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), Gobierno de Aragón and CTP (Work Community of the Pyrenees). We also thank the University of Zaragoza for their financial and material support.

We wish to express our gratitude to the many colleagues 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 will be as successful as this one.

Zaragoza, May, 2013.

The editors.

Contributors

LIST OF

PARTICIPANTS

Abaurrea, Jesús
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
abaurrea@unizar.es

Afere, Benson Ade
Department of Mathematics and Statistics,
Federal Polytechnic, Idah,
Kogi State, Nigeria.
adeafere@yahoo.com

Alberto, Isolina
Departamento de Métodos Estadísticos,
EINA, Universidad de Zaragoza,
María de Luna 3,
50018 Zaragoza, Spain.
isolina@unizar.es

Alcaide, David
Department of Statistics, Operations
Research and Computer Sciences,
University of La Laguna, Spain.
dalcaide@ull.es

Amaya, Jorge
CMM-DIM, Universidad de Chile,
Avda. Blanco Encalada 2120, piso 7,
Santiago de Chile, Chile.
jamaya@dim.uchile.cl

Amo, Mariano
Department of Mathematics,
University of Castilla-La Mancha,
Institute of Applied Mathematics in Science
and Engineering,
Spain.
mariano.amo@uclm.es

Asín, Jesus
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
jasin@unizar.es

Avram, Florin
Department of Mathematics,
University of Pau,
64 000 Pau, France.
florin.avram@univ-pau.fr

Balakrishnan, Narayanaswamy
Department of Mathematics and Statistics,
McMaster University,
Hamilton, Ontario, Canada.
bala@mcmaster.ca

Beamonte, María Asunción
Facultad de Economia y Empresa,
Universidad de Zaragoza,
Zaragoza, Spain.
asunbea@unizar.es

Bobecka, Konstancja
Wydział Matematyki i Nauk Informacyjnych,
Politechnika Warszawska,
Warszawa, Poland.
bobecka@mini.pw.edu.pl

Bordes, Laurent
University of Pau - UMR CNRS 5142,
France.
laurent.bordes@univ-pau.fr

Braekers, Roel
Interuniversity Institute for Biostatistics and
Statistical Bioinformatics,
Universiteit Hasselt,
Agoralaan 1, 3590 Diepenbeek, Belgium.
roel.braekers@uhasselt.be

Casero , Victor Manuel
Escuela Técnica Superior de Ingenieros
Industriales,
Departamento de Matemáticas,
Universidad de Castilla-La Mancha,
Avenida Camilo José Cela 3,
13071 Ciudad Real, Spain.
Victormanuel.Casero@uclm.es

Cebrián, Ana C.
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
acebrian@unizar.es

Collet, Francesca
Departamento de Ciencia e Ingeniería de Materiales e Ingeniería Química,
Universidad Carlos III de Madrid,
Madrid, Spain.
fcollet@ing.uc3m.es

Coolen-Maturi, Tahani
Durham University Business School,
Durham University,
Durham DH1 3LE, UK.
tahani.maturi@durham.ac.uk

Das, Debabrata
Research Scholar,
Indian Institute of Technology,
Bombay, Mumbai, India.
debabrataiitb@gmail.com

de la Rosa de Sa, Sara
Departamento de Estadística e I.O. y D.M.,
Universidad de Oviedo,
33071 Oviedo, Spain.
sara16388@hotmail.com

Doostparast, Mahdi
Department of Statistics,
School of Mathematical Sciences,
Ferdowsi University of Mashhad, Iran.
doostparast@math.um.ac.ir

Ekhosuehi, Virtue U.
Department of Mathematics,
University of Benin,
Benin City, Nigeria
myvectorspace2000@yahoo.com

Ersel, Derya
Department of Statistics,
Hacettepe University,
06800,Beytepe, Ankara, Turkey.
dtektas@hacettepe.edu.tr

Esteban, Luis Mariano
Escuela Universitaria Politécnica de
La Almunia,
Universidad de Zaragoza,
Mayor s/n,
50100 La Almunia de Doña Godina, Spain.
lmeste@unizar.es

Galbete, Arkaitz
Department Estadística e Investigación
Operativa,
University Pública de Navarra,
Campus Arrosadia s/n, 31006,
Pamplona, Spain.
arkaitzgalbete@hotmail.com

Gargallo, Pilar
Facultad de Economia y Empresa,
Universidad de Zaragoza,
Zaragoza, Spain.
pigarga@unizar.es

Gavín, Natalia
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
gs_natali@hotmail.com

Gil, María Ángeles
Departamento de Estadística e I.O. y D.M.,
Universidad de Oviedo,
33071 Oviedo, Spain.
magil@uniovi.es

Gouet, Raúl
Dpto. Ingeniería Matemática and CMM
(UMI 2807, CNRS),
Universidad de Chile,
Av. Blanco Encalada 2120, 837-0459,
Santiago, Chile.
rgouet@dim.uchile.cl

Hadi, Ali S.
Department of Mathematics,
The American University in Cairo,
New Cairo, Egypt.
ahadi@aucegypt.edu

Ishiekwene, Cyril Chukwuka
Department of Mathematics,
Faculty of Physical Sciences,
University of Benin, Nigeria.
cycigar@yahoo.co.uk

Jodrá Esteban, Pedro
Departamento de Métodos Estadísticos,
EINA, Universidad de Zaragoza,
María de Luna 3,
50018 Zaragoza, Spain.
pjodra@unizar.s

Konsowa, Mokhtar
Department of Statistics and Operations Research,
Faculty of Science,
Kuwait University,
Safat 13060, Kuwait
konsowa53@yahoo.com

Koyuncu, Nursel
Department of Statistics,
Hacettepe University,
06800,Beytepe, Ankara, Turkey.
nkoyuncu@hacettepe.edu.tr

Lacruz, Beatriz
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
lacruz@unizar.es

Lahoz, David
Departamento de Métodos Estadísticos,
EINA, Universidad de Zaragoza,
María de Luna 3,
50018 Zaragoza, Spain.
davidla@unizar.es

Leisen, Fabrizio
Departamento de Estadística,
Universidad Carlos III de Madrid,
Madrid, Spain.
fleisen@unav.es

López, F. Javier
Dpto. Métodos Estadísticos and BIFI,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
javier.lopez@unizar.es

López-Fidalgo, Jesús
Department of Mathematics,
University of Castilla-La Mancha,
Institute of Applied Mathematics in Science and Engineering,
Spain.
Jesus.LopezFidalgo@uclm.es

López-Blazquez, Fernando
Dpto. Estadística e I.O.,
Facultad de Matemáticas,
Universidad de Sevilla,
Reina Mercedes, s/n., 41012 Sevilla, Spain.
lopez@us.es

Maldonado, Lina
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
lmguaje@unizar.es

Malina, Magdalena
Mathematical Institute,
Wrocław University,
pl. Grunwaldzki 2/4, 50-385 Wrocław, Poland.
Magdalena.Malina@math.uni.wroc.pl

Mallor, Fermín
Department Estadística e Investigación Operativa,
University Pública de Navarra,
Campus Arrosadia s/n, 31006,
Pamplona, Spain.
mallor@unavarra.es

Martín, Nirian
Department of Statistics,
Carlos III University of Madrid,
Calle Madrid 126,
28903 Getafe (Madrid), Spain.
nirian.martin@uc3m.es

Mata, Raquel
Department of Statistics and O.R. I,
Complutense University of Madrid,
Plaza De Ciencias 3,
28040, Madrid, Spain.
raquel.mata@pdi.ucm.es

Mateo, Pedro
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
mateo@unizar.es

Miranda, Pedro
Complutense University of Madrid,
Madrid, Spain.
pmiranda@mat.ucm.es

Moler, José Antonio
Department Estadística e Investigación Operativa,
University Pública de Navarra,
Campus Arrosadia s/n, 31006,
Pamplona, Spain.
jmoler@unavarra.es

Molina, Manuel
Department of Mathematics,
University of Extremadura, Spain.
mmolina@unex.es

Nafidi, Ahmed
Universit Hassan 1er,
Ecole Suprieure de Technologie- Berrechid,
B.P: 218, Berrechid, Maroc
nafidi@estb.ac.ma

Nuñez, Gabriel
Department of Statistics,
University Carlos III of Madrid,
Madrid, Spain.
gabriel.nunez@uc3m.es

Osagiede, Augustine Aideyan
Department of Mathematics,
University of Benin,
Benin City, Nigeria
austaide2006@yahoo.com

Palacios, Fátima
Departamento de Estadística e I.O.,
Universidad de Sevilla,
Sevilla, Spain.
fatimapalr@hotmail.com

Parast, Layla
Department of Biostatistics,
Harvard School of Public Health,
USA.
lparast@hsph.harvard.edu

Paroissin, Christian
University of Pau - UMR CNRS 5142,
France.
cparoiss@univ-pau.fr

Pérez González, Carlos
Department of Statistics and Operations Research,
University of La Laguna, Spain.
cpgonzal@ull.es

Pérez-Palomares, Ana
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
anapp@unizar.es

Plo, Fernando
Dpto. Métodos Estadísticos,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
fplo@unizar.es

Reyes, Patricio
Department of Statistics,
Carlos III University of Madrid,
Calle Madrid 126,
28903 Getafe (Madrid), Spain.
patricio.reyes.valenzuela@gmail.com

Rivera, María Elena
McGill University,
Burnside Hall, Room 1023,
805 Sherbrooke Street West H3A 2K6,
Montreal QC, Canada.
mancia@math.mcgill.ca

Rodríguez, Licesio J.
Instituto de Matemática Aplicada a la Ciencia y a la Ingeniería,
Universidad de Castilla-La Mancha,
Spain.
L.RodriguezAragon@uclm.es

Rodríguez, Juan Manuel
Department of Statistics,
University of Salamanca,
Spain.
juanmrod@usal.es

Santos, M. Teresa
Department of Statistics,
University of Salamanca,
Spain.
maysam@usal.es

Sanz, Gerardo
Dpto. Métodos Estadísticos and BIFI,
Facultad de Ciencias,
Universidad de Zaragoza,
Pedro Cerbuna, 12, 50009 Zaragoza, Spain.
gerardo@unizar.es

Sicilia, Joaquín
Department of Statistics, Operations Research and Computer Sciences,
University of La Laguna, Spain.
jsicilia@ull.es

Sinova, Beatríz
Departamento de Estadística e I.O. y D.M.,
Universidad de Oviedo,
33071, Oviedo, Spain.
sinovabeatriz.uo@uniovi.es

Turlot, Jean-Christophe
University of Pau - UMR CNRS 5142,
France.
jean-christophe.turlot@univ-pau.fr

Villarroel, Javier
Facultad de Ciencias,
Universidad de Salamanca,
Pza Merced s/n 37008, Salamanca, Spain.
javier@usal.es

Vivo, Juana-María
Department of Quantitative Methods for Economy,
University of Murcia,
Murcia, Spain.
jmvivomo@um.es

Wesolowski, Jacek
Matematyki i Nauk Informacyjnych,
Politechnika Warszawska,
Warszawa, Poland.
wesolo@mini.pw.edu.pl

Zografos, Konstantinos
Department of Mathematics,
University of Ioannina,
451 10 Ioannina, Greece.
kzograf@uoi.gr

Contributed Papers

Monografías Matemáticas García de Galdeano **38**, 23–32 (2013)

APPLICATION OF LOCAL SEARCH TO CREW SCHEDULING

Jorge Amaya, Héctor Ramírez and Paula Uribe

Abstract. This work introduce a model for the crew scheduling problem for train operations, based on a rotative schema, where weekly trips are fixed along the time. This generates a 0-1 medium/large size optimization problem. The special feature of this model is an infinite horizon schedule, due to the rotative schema, where every crew takes the place of the consecutive crew when a new week starts. The problem resolution is performed through three steps: first, finding a feasible solution of infinite length, where schedules repeat in a rotative way between crews; then, an adapted local search is used to improve the initial solution, in order to equilibrate the weekly working hours among crews; finally, drivers are assigned to the scheduled weeks, by solving a flow problem.

Keywords: crew scheduling, integer programming, heuristics.

AMS classification: 90C09, 90C10, 90C27.

§1. Introduction

Crew scheduling is one of the major phases in crew management in large transportation networks such as railway, bus and airline systems, where technical, legal and time constraints must be taken into account when scheduling drivers and crews. A crew in our specific application, typically consists of two drivers, to which a set of tasks (trips) are daily assigned.

Crew assignment (see, for example, [5] and [6]) is a classical optimal decision problem. In general, this assignment problem can have a very high number of decision variables which entails a high degree of complexity for resolution. Frequently, the standard branch and bound strategies are not able to solve large instances, then many variants of well known algorithms have been applied to tackle these hard problems. For a urban bus system, in [3], the authors propose a column generation approach to solve the transit crew scheduling problem. For the air crew rostering problem, in [7], they use a generalized set partitioning model and a method using column generation, adapted to take advantage of the structure of the problem. They claim that this method is capable of solving very large scale problems with thousands of constraints and hundreds of subproblems. An hybrid column generation approach for the urban transit crew problem is studied in [13]. The authors divide the problem in two stages: crew scheduling and crew rostering, solving each separately, and combining mathematical programming and constraint logic programming with column generation. The article cited in [12] describes the development and implementation of an integer optimization model to resolve disruptions to an operating schedule in the rail industry. Favorable results for both the combined train/driver scheduling model and the real-time disruption recovery model are

presented in that paper. Article [1] uses an iterative partitioning for large scale crew scheduling instances; Lagrangean relaxation combined with subgradient optimization is applied in [2]. Decomposition and relaxation strategies are used in [11], for the resolution of a multicommodity network flow problem, representing the railroad crew assignment. Heuristics approaches, such as simulated annealing and genetic algorithms, are proposed in [4], [8] and [9], both for airline and train crews. In [10], the authors apply high performance Integer Optimization for the practical resolution of the crew scheduling problem. They use a Lagrangean relaxation based heuristics and a sequential active set strategy.

The work presented here correspond to a specific application for a Chilean railway company. For this case, the biggest interest is to distribute as balanced as possible the load between crews and to maintain the week load within the legal bounds. The problem resolution must also provide an output composed of a rotative weekly schedule, in which after m weeks, every crew will have met the program of every week. The main advantage of this strategy is to keep a balanced hours load for all crews, besides being an infinite horizon schedule, reusable as many times as desired. The general resolution approach is given in three sequential steps. Firstly, a feasible solution is obtained, which is equivalent to a schedule where every trip is covered, but the working hours load is not necessarily balanced among weeks. this is made by using the Branch and Bound algorithm. Secondly, a local search heuristic is used to improve the initial feasible solution, by balancing the weekly crews load. Finally, crews are assigned to the scheduled weeks, taking into account the initial conditions of crews, in terms of current location, immediately past loads and rest hours.

§2. The conceptual model

The optimization mathematical model can be described through a set of constraints and an objective function, based on the description presented below.

2.1. Constraints

- *One trip, one crew*. Each trip must be assigned to one and only one crew.
- *Legal rest*. For each 7-days window there must be at least 1 legal rest. A legal rest corresponds to a fictitious trip of 33 hours, beginning at 9 PM and ending at 6 AM of the subsequent day.
- *Inter-trips rest*. Between a pair of trips a time window called *inter-trips* rest must be imposed. The duration of that window is given by the labor regulations laws.
- *Sunday rest*. A Sunday rest corresponds to a fictitious trip of 24 hours, beginning at 0:00 hours of Sunday. There are rest regimes of 0, 1 or 2 Sundays rests a month, and it must be assigned according to the specified regime.
- *Origin/Destination*. The origin of a trip must be the destination of the previous one.
- *Consecutive trips*. There are pairs of trips that conform a round trip. In these cases, it is imposed that a trip must be followed by its pair.

- *Rotation*. In a normal schedule, *m* weeks are programmed, in order to assign the work load in a balanced way. Week after week, each crew takes the place of the next one, thus after *m* weeks, every crew will have served the same trips sequence.
- *Fixed rests*. Sometimes, pre-established rests programs are used. The final system must be able of operating either with these programs or allowing the rest days be fixed by the mathematical model.

2.2. Objective Function

Assuming that all trips can be served by a crew, the most critical issue for this application is to schedule as balanced as possible the work load among weeks (or crews).

§3. The mathematical model

Let us denote $V_1, \ldots, V_n$ the set of train trips in a week. We assume that these trips are regular, in the sense that the same scheduling is repeated every week. We include in the set of trips, two sets of virtual trips: the overnight legal and the Sunday rest, that will be explained at the end of this section. Let us denote by $\mathcal{V}$ the set of all trips (including the virtual trips).

Each trip in $\mathcal{V}$ is characterized by a vector of attributes or parameters considered here as given input data for the model. These are: starting time (day, hour, minutes), travel duration, initial station or origin and final station or destination. We also include the *next trip*, which means that a given trip must be followed by another well specified trip, in the special case of round trips. So, we assume that the following information is known:

- N_v is the next trip associated to v;
- I_v and F_v denote the initial station and the final destination, respectively;
- (h_v, m_v) denotes the hour and the minutes of the trip v (then, $0 \leq h_v \leq 24$ and $0 \leq m_v \leq 60$);
- $(\Delta h_v, \Delta m_v)$ denotes the hour and minutes of duration for trip v; and
- $(\bar{h}_v, \bar{m}_v)$ is the arrival time of v.

We consider legal and Sunday rests as **virtual** trips, denoted by v_{LR} and v_{SR}, respectively. An original and simplifying idea in our approach consists in imposing a rotation scheme where a crew i takes the schedule of the crew $i+1$ in the next working week. In this manner, after m weeks (being m the number of crews), all crews take all schedules, which in particular implies that the number of hours done by all the crews are the same in the long term.

We define x_{ivk}, an integer 0-1 variable indicating if crew $i \in \mathcal{T} = \{1, \ldots, m\}$ is allocated to trip $v \in \mathcal{V}$ at day $k \in \mathcal{D} = \{1, \ldots, 7\}$, that is:

$$x_{ivk} = \begin{cases} 1 & \text{if } i \text{ is allocated to trip } v \text{ at day } k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

3.1. Constraints

- *One trip, one crew*. Each real (not virtual) trip can have one and only one crew, so we impose:

$$\sum_{i\in\mathcal{T}} x_{ivk} = 1 \quad \forall v \in \mathcal{V}, k \in \mathcal{D} \tag{2}$$

- *Incompatibility between two trips.* Let us define the compatibility index for a pair of trips: if v, v' are two trips in days k and k', respectively, then we define a parameter $\eta_{vkv'k'}$ by: $\eta_{vkv'k'} = 1$ if (v, k) is compatible with (v', k'), and $\eta_{vkv'k'} = 0$ if not.

  Incompatibility index $\eta_{vkv'k'}$ is calculated considering the time of arrival/departure and the origin/destination of the trips. The incompatibility constraint is then expressed by:

  – For different days $k' > k$:

$$i \in \mathcal{T},\ \ v, v' \in \mathcal{V},\ \ \eta_{vkv'k'} = 0 :\ \ x_{ivk} + x_{iv'k'} \leq 1 \tag{3}$$

  – For the same day $k' = k$:

$$i \in \mathcal{T},\ \ v, v' \in \mathcal{V}, v \neq v',\ \ \eta_{vkv'k'} = 0 :\ \ x_{ivk} + x_{iv'k} \leq 1 \tag{4}$$

  The only exception to this time incompatibility are the virtual trips associated to the rest days of the crews. This means that a *rest trip* is compatible with all the other real trips.

- *Overnight legal rest.* The legal rest must be assigned before the 7^{th} working day, so we impose:

$$\forall v = v_{LR},\ i \in \mathcal{T},\ k \in \mathcal{D} : 1 \leq \sum_{j=k}^{\min(7,k+5)} x_{ivj} + \sum_{j=1}^{k-1} x_{(i+1)vj} \leq 2 \tag{5}$$

- *Sunday rest.* The Sunday rest regime indicated by the R attribute imposes the number of free Sunday in a group of 4 consecutive weeks. The corresponding constraint is written as:

$$\sum_{j=i}^{i+3} x_{jv7} \geq \mathcal{R} \qquad v = v_{SR},\ \ \forall i \in \mathcal{T} \tag{6}$$

- *Crew rotation.* In order to impose that crew i takes the schedule of crew $i + 1$ the next week and so on, we write, for

$i = 1, \ldots, m-1, \forall i \in \mathcal{T},\ \ v, v' \in \mathcal{V},\ \ \eta_{v7v'1} = 0:$

$$x_{iv7} + x_{(i+1)v'1} \leq 1 \tag{7}$$

and, to impose that crew m takes the schedule of crew 1, we write, for $v, v' \in \mathcal{V},\ \eta_{v7v'1} = 0:$

$$x_{mv7} + x_{1v'1} \leq 1 \tag{8}$$

- *Consecutive trips.* If the pair of trips (v, k) and (v', k') are defined as consecutive (served by the same crew), then we impose:

$$x_{ivk} = x_{iv'k'} \quad \forall i \in \mathcal{T},\ \ k \in \mathcal{D} \tag{9}$$

3.2. Objective function

Since the idea is to achieve a balanced weekly amount of working hours for every crew, we define the integer variables z^+ y z^- through the inequality:

$$z^- \leq \sum_{v \in \mathcal{V},\ k \in \mathcal{D}} \Delta_v x_{ivk} \leq z^+ \quad \forall i \in \mathcal{T} \tag{10}$$

where Δ_v is the duration of trip v. So, we use the following objective (balanced hours):

$$\min z^+ - z^- \tag{11}$$

subject to constraints (2)-(10).

§4. The drivers assignment problem

The previous model permits to find a feasible or optimal equilibrated trips diagram, but it doesn't include the identification of crews. For the crew assignment, we propose to consider the previous model as an input, which provides a feasible solution but without identifying the specific crew to be assigned to each weekly diagram.

Let $i \in \mathcal{T}$ given crew and $j \in \mathcal{T}$ be a weekly diagram given by the previous model. We also denote w_{ij} the weight of the crew i to be assigned to week j. This term can be proportional to the difference between the number of hours cumulated by the crew i in the previous week and the number of hours to be done at week j. We use the variable

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ is assigned to week } j \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

This means that each crew is assigned to one and only only one week of the diagram, and each week is assigned to one and only one crew. We also define a bipartite graph whose vertices can be divided into two disjoint sets: the set of crews and the sets of weeks. The set $\mathcal{A}$ of oriented arcs connecting crews to weeks is defined as:

$$(i,j) \in \mathcal{A} \Longleftrightarrow i \text{ is } \textit{compatible} \text{ with } j$$

Compatibility here means that given a crew i, it can be effectively assigned to a given week j. This can be expressed by the following three conditions:

- *Rest hours.* Last service (trip) by crew i must satisfy the minimum rest period with respect to the first trip in the week.
- *Feasible location of the crew.* The actual location of the crew must be equal to the initial location (origin) of the first trip in the scheduled week.
- *Legal rest day.* The last legal rest day of the crew must satisfy the legal rest condition with respect to the scheduled week (on day-off in every 7-days interval).

The objective function of this problem is:

$$\sum_{(i,j)\in\mathcal{A}} w_i y_{ij} \tag{13}$$

which have to be maximized.

Given that the number of weeks of the diagram and the number of available crews are equal, then this problem can be interpreted as to find an optimal one-to-one assignment between crews and weeks. The constraints are:

$$\sum_{i\,/\,(i,j)\in\mathcal{A}} y_{ij} = 1 \quad \forall j \in \mathcal{T} \tag{14}$$

and

$$\sum_{j\,/\,(i,j)\in\mathcal{A}} y_{ij} = 1 \quad \forall i \in \mathcal{T} \tag{15}$$

Expressions (13), (14) and (15) define the optimal assignment of crews to weeks of the diagram.

This is a medium size optimization flow problem whose solution is easy to obtain, in comparison with the computer time for the main scheduling problem given in Section 3.

§5. The adapted local search algorithm

In practice, the problem formulated above is hard to solve, specially due to the constraint (2), which forces to assign every trip to a crew. This complexity can be decreased (in terms of execution time) if the constraint (2) is relaxed as:

$$\sum_{i\in\mathcal{T}} x_{ivk} \leq 1 \quad \forall v \in \mathcal{V}, k \in \mathcal{D} \tag{16}$$

which permits to leave some trips without crew assignment. Then, the optimization problem is now defined by constraints (3.1)-(10) and (16), but with the values z^- and z^+ fixed by the user (the case $z^- = 0$ and $z^+ = \infty$ are also allowed), with the objective function:

$$\max \sum_{i \in \mathcal{T},\ v \in \mathcal{V},\ k \in \mathcal{D}} x_{ivk} \tag{17}$$

which corresponds to maximize the number of served trips. This formulation may leave some uncovered trips (when the original set of constraints is unfeasible), but that can be fixed through the objective function (17).

Given the simplified formulation above, one can solve the problem of finding a balanced trips allocation combining the mathematical model with a heuristic routine which implements local search. The local search routine consists of 3 stages of resolution. First, using an optimization solver we find a feasible solution, where every trip is served, using the relaxed model. The feasible schedule resulting of stage 1 is given as an input for stage 2, where the whole diagram is fragmented into blocks of few weeks (ideally, blocks must have a maximum size 10 weeks). The local search based heuristic is an iterative routine that takes a block, fixes the variables outside it and lets the variables within free for re-optimization, applying the model for finding a balanced solution. This process is repeated m times, travelling through all weeks and solving a sub-problem on each iteration. Finally, one last re-optimization is made, releasing all variables and applying the balanced solution model to the whole diagram, with a time limit constraint in order to ensure the process will end within a reasonable execution time.

This approach takes advantage of the fact that solving a problem using *warm start* strategies decreases the execution time, since the number of feasible branches is immediately reduced in the Branch and Bound algorithm. This, combined with the strong reduction of complexity when multiple sub-problems are solved instead of an unique big problem, highly decreases the execution time and provides very balanced solutions, as we will see in Section 6.

The model (13)-(15), that deals with the assignment of crews to the schedules weeks is a simple bipartite graph, where source nodes are represented by crews and destination nodes by the scheduled weeks. A one-to-one assignment is then performed. The feasibility depends on the initial conditions of crews, mainly the current location, the accumulated worked hours and the last legal rest day. The cost of the arcs is the square of the difference between the normalized coefficients of the crew accumulated load and the load of the scheduled week. Thus, the objective function is to maximize the sum of the arcs pondered by their cost, forcing highly loaded crews be assigned to lightly loaded weeks and vice versa, attempting to maintain a balanced hour schedule after the assignment.

Optimization models and heuristics routines were written in AMPL programming language, that provides enough flexibility for a big range of operations. The routines were packaged within a Java based user interface. The main features of this software is to allow the user to solve different problems for various scenarios, changing parameters such as number of crews, trips attributes and time limit. It also allows the easy interaction for uploading the data files and downloading the output solution, in different format files. The user can remotely submit

big instances of the model using HPC resources.

The interface follows a sequence of stages for each executed instance:

1. *Read/Transform data to AMPL language.*
2. *Connect to the HPC through SSH.*
3. *Send data to High-Performance computer.*
4. *Trigger the execution routine.*

§6. Case studies

In this section, we present numeric results obtained using the algorithm presented in Section 5 for finding a balanced solution, in terms of execution time and performance.

Along the path of this train network, there are different courses that cover various geographic zones with variable extension and operative characteristics. This implies that there are different types of schedules, depending on the operation zone, with variable dimensions in terms of number of variables, according to the number of crews and trips to serve. Through test experiments and models validation, we detected that execution time increases with the number of variables and also, this effect is specially critical when the model for finding a balanced solution is applied.

Tests using the heuristic algorithm shown it is possible to achieve balanced schedules in a third of the time taken by the balanced solution model and this result can be improved even 10 times when the heuristic algorithm is applied to medium size problems. Below, we show some results for large scale and medium size cases, when the executions were run using HPC resources and the licensed optimization solver mentioned before.

The first result corresponds to a complex scenario with 52 crews (weeks) and 33 regular trips (from Monday to Sunday). The problem execution stops at 60.000 seconds (16,7 hours) due to the time limit, set at 60.000 for this case. This solution has a standard deviation of 4,12 for the weekly hours. For the same problem but using the heuristic algorithm, the execution time falls to 22.000 seconds with an optimum solution with 3,06 hours as standard deviation for weekly load.

For a medium size problem, with 35 crews and 20 regular trips, the execution time decreases from 60.000 seconds (detention for time limit), obtained with the balanced hours model, to 1260 seconds when using the heuristic algorithm, while standard deviation for weekly hours goes from 5,03 to 4,98, respectively.

For small size problems with 10 to 15 crews, the balanced hours model works very quickly because the number of variables and constraints of the problem is perfectly handled by the

software, even when the local open source optimization software is used in a common computer. Thus, it doesn't seem very useful for this case the heuristic option, being sometimes more time consuming than the balanced hours model.

§7. Conclusions

We presented a crew scheduling modeling, with the special characteristic of including a rotation constraint that delivers balanced load in terms of work load among crews and also, generates a reusable and infinite time horizon schedule.

The balanced hours model can be slightly modified in order to generate a balanced schedule in a reduced execution time, by relaxing the one-trip one crew constraint and adding upper and lower bounds to the total weekly hours. The risk when we use this model is to obtain infeasible solutions because the constraint of serving all trips is not strict and thus could be violated.

The complexity of the problem when the size of the problem increases leads to high execution times, which was faced by implementing a heuristic algorithm that combines warm start strategies with a local search iterative routine. Results are very encouraging, showing a strong reduction of the execution time for medium and large scale cases.

Acknowledgements

This research has been partially supported by Basal project CMM-Universidad de Chile. The authors also thank FCAB-Antofagasta Railway Co. for complementary support and real data testing.

References

[1] ABBINK, E., J. V. WOUT AND D. HUISMAN. Solving Large Scale Crew Scheduling Problems by using Iterative Partitioning. In: *Proceedings ATMOS 2007, 7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization and Systems*. Ed. by Ch. Liebchen, R. K. Ahuja and J. A. Mesa. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[2] BEASLEY, J. E. AND B. CAO. A tree search algorithm for the crew scheduling problem. *European Journal of Operational Research*, 94 (1996), 517–526.

[3] DESROCHERS, M. AND F. SOUMIS. A Column Generation Approach to the Urban Transit Crew Scheduling Problem. *Transportation Science*, 23 (1989), 1–13.

[4] EMDEN-WEINERT, TH. AND M. PROKSCH. Best Practice Simulated Annealing for the Airline Crew Scheduling Problem . *Journal of Heuristics*, 5 (1999), 419–436.

[5] ERNST, A. T., H. JIANG, M. KRISHNAMOORTHY, H. NOTT AND D. SIER. An Integrated Optimization Model for Train Crew Management. *Annals of Operations Research*, 108 (2001), 211–224.

[6] ERNST, A. T., H. JIANG, M. KRISHNAMOORTHY AND D. SIER. Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, 153 (2004), 3–27.

[7] GAMACHE, M., F. SOUMIS, G. MARQUIS AND J. DESROSIERS. A Column Generation Approach for Large-Scale Aircrew Rostering Problems. *Operations Research*, 47 (1999), 247–263.

[8] JIAN, M-S. AND T-Y. CHOU. Multiobjective genetic algorithm to solve the train crew scheduling problem. In: *I*STASC'10 Proceedings of the 10th WSEAS International Conference on Systems Theory and Scientific Computation, 2010.

[9] LEVINE, D. Application of a hybrid genetic algorithm to airline crew scheduling. *Computers and Operations Research*, 23 (1996), 547–558.

[10] SANDERS, P., T. TAKKULA AND D. WEDELIN. High Performance Integer Optimization for Crew Scheduling. In: 7th International Conference on High Performance Computing and Networking Europe, number 1593 in LNCS, p. 3–12. Springer Verlag, 1999.

[11] VAIDYANATHAN, B., K. C. JHA AND R. K. AHUJA. Multicommodity Network Flow Approach to the Railroad Crew Scheduling Problem. *IBM Journal of Research and Development*, 51 (2007), 325–344.

[12] WALKER, C. G., SNOWDON, J. N. AND RYAN, D. M. . Simultaneous disruption recovery of a train timetable and crew roster in real time. *Computers and Operations Research*, 32 (2005), 2077–2094.

[13] YUNES, T. H., A. V. MOURA AND C. DE SOUZA. Hybrid Column Generation Approaches for Urban Transit Crew Management Problems. *Transportation Science*, 39 (2005), 273–288.

Jorge Amaya
Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático
Universidad de Chile
jamaya@dim.uchile.cl

Héctor Ramírez
Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático
Universidad de Chile
hramirez@dim.uchile.cl

Paula Uribe
Centro de Modelamiento Matemático
Universidad de Chile
pauribe@dim.uchile.cl

Monografías Matemáticas García de Galdeano **38**, 33–42 (2013)

ON SPECTRAL ANALYSIS OF HEAVY-TAILED KOLMOGOROV-PEARSON DIFFUSIONS

Florin Avram, Nikolai N. Leonenko and Nenad Šuvak

Abstract. The self-adjointness of the semigroup generator of one dimensional diffusions implies a spectral representation which has found many useful applications, for example in mathematical finance. However, on non-compact state spaces, the spectrum of the generator will typically include both a discrete and a continuous part, with the latter starting at a spectral cutoff point related to the nonexistence of stationary moments. The significance of this fact for statistical estimation is still not fully understood. We consider here the problem of spectral representation of the transition density for an interesting class of diffusions: the hypergeometric[1] diffusions with heavy-tailed Pearson type invariant distribution, to be called Kolmogorov-Pearson diffusions (Reciprocal (inverse) gamma, Fisher-Snedecor and skew-Student diffusions). As opposed to the "classic" hypergeometric diffusions (Ornstein-Uhlebeck, Gamma/CIR, Beta/Jacobi), these diffusions have a continuous part of the spectrum, whose spectral cutoff and transition density we provide in an explicit form.

Keywords: Diffusion process, Infinitesimal generator, Kolmogorov-Pearson diffusion, Sturm-Liouville equation, Transition density.

AMS classification: 33C47, 60G10, 60J25, 60J60, 62M15.

§1. Introduction

We study here some analytical and probabilistic properties of diffusion processes

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt \tag{1}$$

with quadratic diffusion coefficient $a(x) = \frac{\sigma^2(x)}{2} = \epsilon(a_2x^2 + a_1x + a_0)$, $\epsilon > 0$, $a_2 > 0$ and linear drift $b(x) = -\theta(x - \mu)$. We may write (1) as

$$dX_t = \theta(\mu - X_t)\, dt + \sqrt{2\epsilon a_2 \left[(X_t - \mu')^2 + \delta^2\right]}\, dB_t, \quad t \geq 0. \tag{2}$$

This parametrization, to be called Student parametrization, makes sense for the whole Kolmogorov-Pearson family (by allowing $\delta^2 \leq 0$), but it is especially convenient when $\delta \in \mathbb{R}, a_2 > 0$, in which case it produces diffusions living on $(-\infty, \infty)$.

NAME	$a(x)$	SPEED DENSITY
Ornstein-Uhlenbeck	$1/2$	$e^{-\theta(x-\mu)^2}$
CIR/Squared O-U/Gamma	x	$x^{p-1}e^{-\theta x}$
Jacobi/Beta	$x(a_1-a_2x), a_2>0, \Delta(a)>0$	$x^{\frac{p}{a_1}-1}(a_1-a_2x)^{-\frac{p}{a_1}+\frac{\theta}{a_2}-1}$
Reciprocal gamma /IGBM	$a_2x^2, a_2>0, \Delta(a)=0$	$x^{-\frac{\theta}{a_2}-2}e^{-\frac{p}{a_2x}}$
Student/hypergeometric	$a_2x^2+a_0, a_2>0, \Delta(a)<0$	$\sigma(x)^{-\frac{\theta}{a_2}-2}\, e^{\frac{p}{\sqrt{a_0a_2}}\tan^{-1}\left(\frac{x}{\sqrt{a_0/a_2}}\right)}$
Fisher-Snedecor	$x(a_1+a_2x), a_2>0, \Delta(a)>0$	$x^{\frac{p}{a_1}-1}(a_1+a_2x)^{-\frac{p}{a_1}-\frac{\theta}{a_2}-1}$

Table 1: The speed density for Kolmogorov-Pearson diffusions

A classification of Kolmogorov-Pearson diffusions may be achieved by using the degree of the polynomial $a(x)$ from the diffusion coefficient, the sign of its leading coefficient a_2 and the discriminant $\Delta(a)$ in the quadratic case (see Table 1).

The Ornstein-Uhlenbeck, CIR/Gamma and Jacobi/Beta diffusions, that have all moments and complete orthogonal polynomial bases, have been extensively studied and widely applied (especially in the ergodic case). However, the first results on the statistical analysis of the heavy-tailed diffusions (1) are more recent. We study these processes under the assumption of non-regular boundaries, which ensures the uniqueness of the diffusion. Furthermore, under this assumption the specification of boundary conditions is not required. For more detailed study of heavy-tailed Kolmogorov-Pearson diffusions we refer to the complete version of the paper, see [3].

§2. Heavy-tailed Kolmogorov-Pearson diffusions

We observe the class of diffusion processes defined by the SDE (1) with the linear drift $b(x)=b_1x+b_0$ and the quadratic squared diffusion coefficient

$$\sigma^2(x)=2a(x)=2(a_2x^2+a_1x+a_0)=\sigma_2^2x^2-\sigma_1^2x+\sigma_0^2.$$

Heavy-tailed Kolmogorov-Pearson diffusions with the state space $\langle l,r\rangle$ are a subclass of this class of diffusions and are characterized by properties given in Table 1. Existence of the unique Markovian weak solution $X=\{X_t,\, t\geq 0\}$ of the SDE (1) with the pre-specified marginal density from the Pearson family follows from Bibby et. al. [4, Theorem 2.1.(i)]. Furthermore, the SDE (1) admits a unique strong solution with the time-homogenous transition densities if it satisfies the following sufficient conditions given by Aït-Sahalia [1, page 415, assumption A1 (i) and (ii)]:

- the drift coefficient $b(x)$ and the diffusion coefficient $\sigma(x)$ are continuously differentiable in x and $\sigma^2(x)$ is strictly positive on the whole diffusion state space,
- the integral of the speed density $\mathbf{m}(x)$ of diffusion X converges at both boundaries of the diffusion state space.

[1] It seems appropriate to call this class of processes hypergeometric diffusions, due to the appearance of the Gauss ${}_2F_1$ function and its limiting confluent forms in various explicit formulas.

According to Aït-Sahalia [1], these conditions are considerably less restrictive than the global Lipschitz and the linear growth conditions. Existence of strong solutions for particular Kolmogorov-Pearson diffusions is verified in [2, Section 3], [8, Section 3] and [9, Section 3], respectively.

2.1. Oscillatory/non-oscillatory classification of natural boundaries

The infinitesimal generator $\mathcal{G}f(x) = a(x)f''(x) + b(x)f'(x)$ plays the crucial role in classification of boundaries of the diffusion state space. For O/NO classification of the natural boundaries $l = -\infty$ and $r = \infty$ of Kolmogorov-Pearson diffusion, the standard procedure requires transformation of the Sturm-Liouville equation

$$a(x)f''(x) + b(x)f'(x) + \lambda f(x) = 0, \quad \lambda \geq 0 \tag{3}$$

to the Liouville normal form (Fulton et al. [6, pg. 4])

$$-g''(u) + Q(u)g(u) = \lambda g(u). \tag{4}$$

The function $Q(u)$ is called the potential function and in the case of the diffusion (1) is given by

$$Q(u) = \frac{e^{4\sqrt{a_2}u}(a_2 - b_1)^2}{4a_2\left(e^{2\sqrt{a_2}\,u} - \Delta(a)\right)^2} - \frac{e^{3\sqrt{a_2}\,u}(b_1 - 2a_2)(a_1b_1 - 2a_2b_0)}{a_2\left(e^{2\sqrt{a_2}\,u} - \Delta(a)\right)^2} -$$
$$- \frac{4e^{\sqrt{a_2}\,u}(b_1 - 2a_2)(a_1b_1 - 2a_2b_0)\Delta(a) - (a_2 - b_1)^2\Delta^2(a)}{4a_2\left(4e^{2\sqrt{a_2}\,u} - \Delta(a)\right)^2} +$$
$$+ \frac{e^{2\sqrt{a_2}u}\left(-8a_1a_2b_0b_1 + a_1^2\left(5a_2^2 - 6a_2b_1 + 3b_1^2\right) - 4a_2\left(-2a_2b_0^2 + a_0\left(5a_2^2 - 6a_2b_1 + b_1^2\right)\right)\right)}{2a_2\left(e^{2\sqrt{a_2}\,u} - \Delta(a)\right)^2}.$$

Natural boundaries $l = -\infty$ and $r = \infty$ of the Sturm-Liouville equation (3) remain unchanged under the Liouville transform, i.e. the corresponding boundaries of the equation (4) are $l^* = u(l) = -\infty$ and $r^* = u(r) = \infty$. O/NO classification of the boundaries $l^* = -\infty$ and $r^* = \infty$ depends on the behavior of the potential function $Q(u)$ near these boundaries. Since the last three terms in the expression for $Q(u)$ vanish as $u \to -\infty$ and $u \to \infty$, it follows that

$$\lim_{u\to-\infty} Q(u) = \lim_{u\to\infty} Q(u) = \frac{(a_2 - b_1)^2}{4a_2}.$$

By using [6, Theorem 6], we have shown that both $l^* = -\infty$ and $r^* = \infty$ are O/NO boundaries of the equation (4) with unique positive cutoff

$$\Lambda = \Lambda(a_2, b_1) = \frac{(a_2 - b_1)^2}{4a_2}. \tag{5}$$

Furthermore, we verified that these boundaries are NO for $\lambda < \Lambda$ and O for $\lambda > \Lambda$. According to Dunford and Schwartz [5, Corollary 57, pg. 1481] (see also Linetsky [10, Theorem

3, pg. 349]), both boundaries are NO for $\lambda = \Lambda$. This classification of boundaries remains invariant under the Liouville transform (see [6, pg. 6]), i.e. both $l = -\infty$ and $r = \infty$ are O/NO boundaries of the Sturm-Liouville equation (3) with unique positive cutoff

$$\Lambda = \Lambda(a_2, b_1) = \frac{(a_2 - b_1)^2}{4a_2}.$$

Furthermore, both boundaries are NO for $\lambda \leq \Lambda$ and O for $\lambda > \Lambda$.

For Reciprocal gamma and Fisher-Snedecor diffusions the left boundary of the diffusion state space is $l = 0$ and it is regular, entrance or exit, depending on parameter values (see [2, 8, 9]). For both diffusions the right boundary $r = \infty$ is natural. Since the cutoff between the discrete and the absolutely continuous spectrum is determined by asymptotic behavior of the potential function near natural boundaries it means that the cutoff formula (5) holds also for these two diffusions living on $\langle 0, \infty \rangle$.

2.2. Spectrum of the Sturm-Liouville operator

Explicit form of the spectral representation of the transition density of the diffusion process is implied by the structure of the spectrum of the corresponding Sturm-Liouvuille operator $(-\mathcal{G})$. Furthermore, according to [10, pg. 350], if the corresponding potential function $Q(x)$ has bounded variation on some subinterval $\langle c, \infty \rangle$ of the positive halfline, in the continuous part of the spectrum of the operator $(-\mathcal{G})$ with natural boundary $r = \infty$ there are no gaps containing simple eigenvalues.

Spectral category and the structure of the spectrum of the heavy-tailed Kolmogorov-Pearson diffusions is given below, according to the general results on O/NO classification of boundaries of the diffusion state space $\langle l, r \rangle$ (see [6, pg. 23-27] and [10, pg. 348, Theorem 2]):

- **l NO boundary, r O/NO natural boundary (Reciprocal gamma and Fisher-Snedecor diffusions)**
  These diffusions belong to Linetsky's spectral category II. In particular, $l = 0$ is NO boundary, while $r = \infty$ is O/NO boundary with unique cutoff

  $$\Lambda = \Lambda(a_2, b_1) = \frac{(a_2 - b_1)^2}{4a_2}. \tag{6}$$

  Since $r = \infty$ is NO for $\lambda = \Lambda$, the Sturm-Liouville operator $(-\mathcal{G})$ has a finite set of simple eigenvalues in $[0, \Lambda]$ and an essential spectrum $\sigma_e(-\mathcal{G}) = [\Lambda, \infty\rangle$. Hence, the operator $(-\mathcal{G})$ has a discrete spectrum $\sigma_d(-\mathcal{G})$ in $[0, \infty\rangle$, i.e. $\sigma_d(-\mathcal{G}) \subset [0, \Lambda\rangle$, and a purely absolutely continuous spectrum $\sigma_{ac}(-\mathcal{G})$ of multiplicity one in $\langle \Lambda, \infty\rangle$. For more details on boundary classification for these two diffusions see [8] and [2].

- **l and r are natural O/NO boundaries (Student diffusion)**
  These diffusions belong to Linetsky's spectral category III. In particular, the Sturm-Liouville operator $(-\mathcal{G})$ has a finite set of simple eigenvalues in $[0, \Lambda]$ and an essential spectrum $\sigma_e(-\mathcal{G}) = [\Lambda, \infty\rangle$. Hence, the operator $(-\mathcal{G})$ has a discrete spectrum $\sigma_d(-\mathcal{G})$ in $[0, \infty\rangle$, i.e. $\sigma_d(-\mathcal{G}) \subset [0, \Lambda\rangle$, and a purely absolutely continuous spectrum

$\sigma_{ac}(-\mathcal{G})$ of multiplicity two in $\langle \Lambda, \infty \rangle$. For more details on boundary classification for symmetric case of Student diffusion see [9].

2.2.1. Discrete part of the spectrum

The discrete part of the spectrum of the operator $(-\mathcal{G})$ is of the form $\sigma_d(-\mathcal{G}) = \{\lambda_n,\, n = 0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}$. Eigenvalues λ_n are given by the explicit expression

$$\lambda_n = n\left((1-n)a_2 - b_1\right), \quad n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}. \tag{7}$$

Corresponding eigenfunctions are polynomial solutions of the Sturm-Liouville equation (3) given by the Rodrigues formula

$$\widetilde{P}_n(x) = \frac{1}{\mathbf{m}(x)} \frac{d^n}{dx^n} \left\{2^n a^n(x) b(x)\right\}, \quad n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}. \tag{8}$$

Polynomials $\widetilde{P}_n(x)$ form the finite system of polynomials orthogonal with respect to the speed density $\mathbf{m}(x)$, i.e.

$$\int_{-\infty}^{\infty} \widetilde{P}_m(x)\, \widetilde{P}_n(x)\, \mathbf{m}(x)\, dx = 0, \qquad m, n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}, \quad m \neq n.$$

For normalization of polynomials with respect to the speed density we must multiply each of them by the normalizing constant given by the general expression

$$K_n = \frac{(-1)^n}{\sqrt{(-1)^n\, n!\, d_n\, I_n}}, \quad n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\},$$

where

$$d_n = 2^n a_2^{(n-1)} (b_1 + (n-1)a_2) \frac{\Gamma\left(\frac{b_1}{a_2} + 2n - 1\right)}{\Gamma\left(\frac{b_1}{a_2} + n\right)}, \qquad I_n = 2^n \int_l^r a^n(x) \mathbf{m}(x)\, dx.$$

Therefore, the normalized orthogonal polynomials are given by the Rodrigues formula

$$P_n(x) = K_n\, \widetilde{P}_n(x), \quad n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}. \tag{9}$$

Remark 1. Orthogonality relation for the normalized polynomials $P_n(x)$, i.e. the relation

$$\int_l^r P_m(x) P_n(x) \mathbf{m}(x)\, dx = \delta_{mn}, \quad m, n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}, \tag{10}$$

implies interesting properties of the random variables $P_n(X_t)$, where X_t is from the diffusion with the state space $\langle l, r \rangle$ and the speed density $\mathbf{m}(x)$. In particular, random variables $P_n(X_t)$ are orthonormal, i.e.

$$E[P_m(X_t)\, P_n(X_t)] = \int_l^r P_m(x) P_n(x) \mathbf{m}(x)\, dx = \delta_{mn}, \quad m, n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}.$$

Since $P_0(x) = 1$, for $m = 0$ and $n \neq 0$ the previous expression takes the form

$$E[P_n(X_t)] = 0, \quad n \in \{1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\},$$

i.e. $P_n(X_t)$, $n = 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor$, are orthonormal centered random variables.

2.2.2. Essential part of the spectrum

In the subsection 2.2 we verified that the essential spectrum of the Sturm-Liouville operator $(-\mathcal{G})$ is $\sigma_e(-\mathcal{G}) = [\Lambda, \infty\rangle$. Moreover, the operator $(-\mathcal{G})$ has purely absolutely continuous spectrum of multiplicity one in $\langle \Lambda, \infty\rangle$, where

$$\Lambda = \frac{(a_2 - b_1)^2}{4a_2},$$

is the unique positive cutoff between the discrete and the absolutely continuous part of the spectrum.

Remark 2. Polynomial eigenfunctions $P_n(x)$, $n \in \{0, 1, \ldots, \lfloor (a_2 - b_1)/2a_2 \rfloor\}$, and (eigen) functions related to absolutely continuous spectrum (i.e. for $\lambda > \Lambda$) belong to orthogonal subspaces of the Hilbert space $\mathcal{H} = L^2(\langle l, r\rangle, \mathbf{m}(\cdot))$. Polynomials $P_n(x)$ belong to the subspace $\mathcal{H}_{pp}$ of the Hilbert space $L^2(\langle -\infty, \infty\rangle, \mathbf{m}(\cdot))$ containing functions having only the pure point spectral measure. Functions related to absolutely continuous spectrum belong to the subspace $\mathcal{H}_{ac}$ of the same Hilbert space, that contains functions having only the spectral measure which is absolutely continuous with respect to the Lebesgue measure (see Linetsky [10, Appendix]). From these facts it follows that polynomial eigenfunctions and functions related to absolutely continuous spectrum are orthogonal with respect to the density $\mathbf{m}(\cdot)$.

2.3. Student diffusion

In this section we apply some results by Shaw [12] and techniques by Paulsen [11, Theorem A.1,pg. 984]. Namely, the Student parametrization of Kolmogorov-Pearson diffusions could be derived by following the formulation of [12]. Consider the SDE

$$\begin{aligned} dX_t &= \theta\,(\mu - X_t)\; dt + \sigma_1\, dW_t^{(1)} + \sigma_2\, X_t dW_t^{(2)}, \quad t \geq 0, \quad \Rightarrow \\ dX_t &= \theta\,(\mu - X_t)\; dt + \sqrt{\sigma_2^2 \left(\left(X_t + \rho \frac{\sigma_1}{\sigma_2} \right)^2 + (1 - \rho^2) \left(\frac{\sigma_1}{\sigma_2} \right)^2 \right)}\, dW_t, \end{aligned} \tag{11}$$

where $W_t^{(1)}$ and $W_t^{(2)}$, $t \geq 0$, are standard Brownian motions with correlation ρ, and $\{W_t,\, t \geq 0\}$ is a standard Brownian motion resulting from combining the two. This diffusion process (11) is Markovian, with infinitesimal generator

$$\mathcal{G}h(x) \quad = \quad a(x)h''(x) + b(x)h'(x) \tag{12}$$

where $2a(x) = \sigma^2(x) = \sigma_2^2 \left(\left(X_t + \rho \frac{\sigma_1}{\sigma_2} \right)^2 + (1 - \rho^2) \left(\frac{\sigma_1}{\sigma_2} \right)^2 \right)$.

Simplifying the notation, we will consider the stochastic differential equation (SDE)

$$dX_t = -\theta(X_t - \mu)\,dt + \sqrt{2a_2\left((X_t-\mu')^2+\delta^2\right)}\,dW_t, \quad t \geq 0, \tag{13}$$

where $\delta > 0$, $2a_2 = \sigma_2^2 = \frac{2\theta}{\nu-1} > 0$, $\mu, \mu' \in \mathbb{R} > 0$, and $W = \{W_t, t \geq 0\}$ is a standard Brownian motion. Note that by a change of origin, one may always assume that either μ or μ' are 0. Putting $\tilde{X}_t := (X_t - \mu')/\delta$, we arrive at

$$d\tilde{X}_t = -\theta(\tilde{X}_t - \tilde{\mu})\,dt + \sqrt{2a_2\left(1+\tilde{X}_t^2\right)}\,dW_t, \quad t \geq 0, \tag{14}$$

where we put $\tilde{\mu} := (\mu - \mu')/\delta$. The infinitesimal operator is

$$\mathcal{G} \;=\; a_2\left(1+\tilde{x}^2\right)D_{\tilde{x}}^2 - \theta(\tilde{x}-\tilde{\mu})D_{\tilde{x}}, \tag{15}$$

and the scale and speed densities are:

$$\mathfrak{s}(\tilde{x}) = (\tilde{x}^2+1)^{\frac{1}{2\tilde{a}}}\; e^{-\frac{\tilde{\mu}}{\tilde{a}}\,\mathrm{arctg}\,(\tilde{x})}, \quad \mathbf{m}(\tilde{x}) = \frac{e^{\frac{\tilde{\mu}}{\tilde{a}}\;\mathrm{arctg}\,(\tilde{x})}}{(\tilde{x}^2+1)^{\frac{1}{2\tilde{a}}+1}}, x \in \mathbb{R},$$

where $\tilde{a} = a_2/\theta$.

A direct approach for solving the corresponding Sturm-Liouville equation in this case was provided by Paulsen [11, Theorem A.1, pg. 984], who finds that the monotone solutions are given by the Weyl type fractional integrals. We revisit now Paulsen' approach (see [11, Thm A1] and also[7]) to the Student Sturm-Liouville operator, which is based on representing the solution as a Weyl fractional integral with kernel $K(x,t) = (t-x)^{n-1+\rho}$. This approach uses Euler's transformation [7, 8.31] which decomposes the original operator in a sum $G = \Gamma_0 - \Gamma_1 + ...(-1)^p\Gamma_p$ of special type operators, involving an additional parameter ρ chosen to minimize p. For the normalized Student SL equation [11, (A2)]

$$G = (x^2+1)D^2 + (\tilde{r}x + \tilde{p}\delta)D - \tilde{\lambda},$$

where tilde signifies the corresponding coefficient is divided by a_2, Euler's decomposition [7, VIII.31] is $G = \Gamma_0 - \Gamma_1 + \Gamma_2$ where

$$\begin{aligned}
\Gamma_0 &= G_0(x)D^2 - \rho G_0'(x)D + \frac{\rho_{(2)}}{2}G_0''(x) = (x^2+1)D^2 - 2x\rho D + \rho(\rho+1)\\
\Gamma_1 &= G_1(x)D - (\rho+1)G_1'(x) = -(\tilde{r}x + \delta\tilde{p} + 2x\rho)D + (\rho+1)(\tilde{r}+2\rho),\\
\Gamma_2 &= (\rho+1)(\tilde{r}+2\rho) - (\rho+1)(\rho) - \tilde{\lambda} = (\rho+1)(\tilde{r}+\rho) - \tilde{\lambda}.
\end{aligned}$$

The last operator vanishes when $\rho^2 + \rho(1+\tilde{r}) + \tilde{r} - \lambda = 0$, for

$$\rho = \frac{1}{2}\left(-1-\tilde{r} \pm \sqrt{(1+\tilde{r})^2 - 4(\tilde{r}-\tilde{\lambda})}\right). \tag{16}$$

In order to stress out the dependence of ρ on the spectral parameter λ we will denote ρ by $\rho(\lambda)$. With these choices, the order p of $M_z = \sum_{i=0}^{p}\Gamma_i(z)D_z^{p-i}$ becomes $p = 1$, i.e.

$M_z = \Gamma_0(z)D_z + \Gamma_1(z)$, and we can apply directly the results of Pochhammer and Jordan [7, XVIII.4], yielding

$$\phi_{\rho(\lambda)}(x) = \int_\Gamma (z-x)^{\rho(\lambda)+1} K(z) dz,$$

where

$$K(z) = \frac{1}{G_0(z)} e^{\int^z \frac{G_1(y)}{G_0(y)} dy} = (z^2+1)^{-1} e^{-\int^z \frac{\delta\tilde{p}+\beta y}{y^2+1} dy} = (z^2+1)^{-\beta/2-1} e^{-\delta\tilde{p}\,\mathrm{arctg}(z)}, \quad (17)$$

$$\beta = \tilde{r} + 2\rho(\lambda). \quad (18)$$

Finally, the contour must be chosen so that $\int d[(z-x)^{\rho(\lambda)}(z^2+1)^{-\beta/2} e^{-\delta\tilde{p}\,\mathrm{arctg}(z)}]$ is 0. If $\rho(\lambda)$ is chosen as the positive root, x is a zero of the total differential, and it may be checked by limiting arguments that the Weyl type fractional integrals

$$\phi^{(1)}_{\rho(\lambda)}(x) = \int_x^\infty (z-x)^{\rho(\lambda)+1} K(z)\, dz, \quad (19)$$

$$\phi^{(2)}_{\rho(\lambda)}(x) = \int_{-\infty}^x (x-z)^{\rho(\lambda)+1} K(z)\, dz, \quad (20)$$

are convergent and that the bilinear concomitant $(z-x)^{\rho(\lambda)}(z^2+1)^{-\beta/2} e^{-\delta\tilde{p}\,\mathrm{arctg}(z)}$ is zero at ∞, certifying thus $\phi^{(1)}_{\rho(\lambda)}(x), \phi^{(2)}_{\rho(\lambda)}(x)$ as two basic solutions of our Sturm-Liouville equation. It is easy to check that:

$$(\phi^+_\lambda)'(y) = -\rho\phi^+_{\lambda-1}(y),$$
$$(\phi^-_\lambda)'(y) = \rho\phi^-_{\lambda-1}(y),$$

where

$$-1 < Re(\rho+1) < 1 + Re(\beta).$$

These may be furthermore checked to be precisely the increasing and decreasing Sturm-Liouville solutions $\phi^+_{\rho(\lambda)}(x)$ and $\phi^-_{\rho(\lambda)}(x)$, respectively, with the Wronskian

$$W_{\rho(\lambda)} = \left(\phi^+_{\rho(\lambda)}(x)\right)' \phi^-_{\rho(\lambda)}(x) - \phi^+_{\rho(\lambda)}(x) \left(\phi^-_{\rho(\lambda)}(x)\right)' =$$

$$= -\rho(\lambda) \left(\phi^+_{\rho(\lambda)-1}(x)\phi^-_{\rho(\lambda)}(x) + \phi^+_{\rho(\lambda)}(x)\phi^-_{\rho(\lambda)-1}(x)\right) =$$

$$= \rho(\lambda) \left(\int_x^\infty (z-x)^{\rho(\lambda)} K(z) dz \int_{-\infty}^x (x-z)^{\rho(\lambda)+1} K(z) dz + \right.$$

$$\left. + \int_x^\infty (z-x)^{\rho(\lambda)+1} K(z) dz \int_{-\infty}^x (x-z)^{\rho(\lambda)} K(z) dz \right).$$

It is well known that, once the Green function is known, its residues and values along an eventual branch cut determine the spectral expansion of the transition density. However,

completing all the details of this complex analysis exercise is quite tedious. For spectral representation of the transition density of positive recurrent Reciprocal gamma and Fisher-Snedecor diffusions with non-regular boundaries we refer to [8], [9] and [2]. For regular boundaries several cases need to be analyzed.

The essential feature of the heavy-tailed case is the presence of a branch cut in the Green's function which, beside the discrete spectrum, produces a continuous part of the spectrum of the corresponding infinitesimal generator. This is in contrast with with the Ornstein-Uhlenbeck, CIR and Jacobi diffusions having only purely discrete simple spectrum and a complete set of (classical) orthogonal polynomials eigenfunctions (Hermite, Laguerre and Jacobi, respectively).

2.4. Spectral representation of transition density

In this section we give the explicit formulae for the discrete and the continuous part of the spectral representation of the transition density of heavy-tailed Kolmogorov-Pearson diffusions.

Eigenvalues λ_n given by (7) are precisely simple poles of the Green's function (they appear in the inverse Laplace transformation of the Green's function). Similarly to spectral representation of the transition density for the regular Sturm-Liouville problems, it follows that the discrete part of the spectral representation of the transition density for heavy-tailed Kolmogorov-Pearson diffusions is of the form

$$p_d(x;y,t) = \sum_{n=0}^{\lfloor (a_2-b_1)/2a_2 \rfloor} e^{-\lambda_n t} P_n(x) P_n(y) \mathbf{m}(x),$$

where $\mathbf{m}(\cdot)$ is the speed density. The non-normalized continuous part of the spectral representation of the transition density for heavy-tailed Kolmogorov-Pearson diffusions is of one of the following forms, depending on spectral category of diffusion:

- Diffusions belonging to the spectral category II (Reciprocal gamma and Fisher-Snedecor diffusion):
$$p_c(x;y,t) = \mathbf{m}(x) \int_\Lambda^\infty e^{-\lambda t} \phi^+_{\rho(\lambda)}(x) \phi^+_{\rho(\lambda)}(y)\, d\lambda,$$
where the increasing functions $\phi^+_{\rho(\lambda)}(\cdot)$ for Reciprocal gamma and Fisher-Snedecor diffusions are given in Leonenko and Šuvak [8, 9].

- According to Linetsky [10], for diffusions belonging to spectral category III (Student diffusion) spectral representation of transition density is of the following form:
$$p_c(x;y,t) = \mathbf{m}(x) \int_\Lambda^\infty e^{-\lambda t} \Big(\phi^+_{\rho(\lambda)}(x) \phi^+_{\rho(\lambda)}(y) + \phi^+_{\rho(\lambda)}(x) \phi^-_{\rho(\lambda)}(y) +$$
$$+ \phi^+_{\rho(\lambda)}(y) \phi^-_{\rho(\lambda)}(x) + \phi^-_{\rho(\lambda)}(x) \phi^-_{\rho(\lambda)}(y) \Big)\, d\lambda,$$
where monotone solutions $\phi^+_{\rho(\lambda)}(\cdot)$ and $\phi^-_{\rho(\lambda)}(\cdot)$ are given by (19) and (20), respectively.

For explicit expressions for spectral representations of transition densities of Reciprocal gamma diffusion, special case of Student diffusion (symmetric Student diffusion) and Fisher-Snedecor diffusion we refer to [8, Theorem 3.1.], [9, Section 4.5.] and [2, Theorem 4.1.], respectively.

References

[1] AÏT-SAHALIA, Y. Testing continuous time models of the spot interest rate. *Rev. Financ. Stud.* ***9*** (1996), 385–426.

[2] AVRAM, F., LEONENKO, N.N. AND ŠUVAK, N. Spectral representation of transition density of Fisher-Snedecor diffusion. *Stochastics: An International Journal of Probability and Stochastic Processes, published on line 11/03/2013, DOI:10.1080/17442508.2013.775285. Extended Arxiv preprint arXiv:1007.4909* (2013).

[3] AVRAM, F., LEONENKO, N.N. AND ŠUVAK, N. On spectral analysis of heavy-tailed Kolmogorov-Pearson diffusions. *Markov Processes and Related Fields, accepted for publication* (2013).

[4] BIBBY, B.M., SKOVGAARD, I.M. AND SØRENSEN, M. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* ***11*** (2005), 281–299.

[5] DUNFORD, N. AND SCHWARTZ, S. *Linear Operators. Part II: Spectral Theory (Self-Adjoint Operators in Hilbert Spaces*. Wiley, city, 1963.

[6] FULTON, C.T., PRUESS, S. AND XIE, Y. The automatic classification of Sturm-Liouville problems. *J. Appl. Math. Comput.* ***124*** (2005), 149–186.

[7] INCE, E.L. *Ordinary Differential Equations*. Dover, city, 1956.

[8] LEONENKO, N.N. AND ŠUVAK, N. Statistical inference for reciprocal gamma diffusion process. *J. Statist. Plann. Inference* ***140*** (2010), 30–51.

[9] LEONENKO, N.N. AND ŠUVAK, N. Statistical inference for Student diffusion process. *Stoch. Anal. Appl.* ***28*** (2010), 972–1002.

[10] LINETSKY, V. The spectral decomposition of the option value. *Int. J. Theor. Appl. Finance* ***7*** (2004), 337–384.

[11] PAULSEN, J. AND GJESSING, H.K. Ruin theory with stochastic return on investments. *Adv. in Appl. Probab.* ***29*** (1997), 965–985.

[12] SHAW, W.T. A model of returns for the post-credit-crunch reality: Hybrid Brownian motion with price feedback. *Arxiv preprint arXiv:0811.0182* (2008).

F. Avram
Department of Mathematics, University of Pau, 64 000 Pau, France
florin.avram@univ-pau.fr

N.N. Leonenko
School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF244AG, UK
LeonenkoN@Cardiff.ac.uk

N. Šuvak
Department of Mathematics, University of Osijek, Trg ljudevita Gaja 6, 31 000 Osijek, Croatia
nsuvak@mathos.hr

Monografías Matemáticas García de Galdeano **38**, 45–53 (2013)

ON PAIRWISE COMPARISON WITH COMPETING RISKS

Tahani Coolen-Maturi

Abstract. In reliability, failure data often correspond to competing risks, where several failure modes can cause a unit to fail. This paper presents nonparametric predictive inference (NPI) for pairwise comparison with competing risks data, assuming that the failure modes are independent. These failure modes could be the same or different among the two groups, and these can be both observed and unobserved failure modes. NPI is a statistical approach based on few assumptions, with inferences strongly based on data and with uncertainty quantified via lower and upper probabilities. The focus is on the lower and upper probabilities for the event that the lifetime of a future unit from one group, say Y, is greater than the lifetime of a future unit from the second group, say X. Finally, an example is given for illustration purposes.

Keywords: Competing risks, reliability, pairwise comparison, nonparametric predictive inference, lower and upper probabilities, lower and upper survival functions, right-censored data.

AMS classification: 62G86, 62N05, 62N99

§1. Introduction

In reliability, failure data often correspond to competing risks [2, 18, 19], where several failure modes can cause a unit to fail, and where failure occurs due to the first failure event caused by one of the failure modes. Throughout this paper, it is assumed that each unit cannot fail more than once and it is not used any further once it has failed, and that a failure is caused by a single failure mode which, upon observing a failure, is known with certainty. Also we assume throughout that the failure modes are independent, inclusion of assumed dependence would be an interesting topic for future research, but cannot be learned about from the data as considered here as shown by Tsiatis [20].

Comparison of two groups or treatments with competing risks is a common problem in practice. For example in medical applications, one may want to compare two treatments with multiple competing risks [15], or in reliability one may want to study the effect of the brand of air-conditioning systems which can fail either due to leaks of refrigerant or wear of drive belts [17].

In this paper we introduce nonparametric predictive inference (NPI) for comparison of two groups with competing risks. NPI is a statistical method based on Hill's assumption $A_{(n)}$ [13], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities [1, 3]. $A_{(n)}$ does not assume anything else, and can be interpreted as a post-data assumption related to exchangeability

[12], a detailed discussion of $A_{(n)}$ is provided by Hill [14]. Inferences based on $A_{(n)}$ are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides bounds for probabilities via the 'fundamental theorem of probability' [12], which are lower and upper probabilities in interval probability theory [1, 22, 23, 24].

In reliability and survival analysis, data on event times are often affected by right-censoring, where for a specific unit or individual it is only known that the event has not yet taken place at a specific time. Coolen and Yan [8] presented a generalization of $A_{(n)}$, called 'right-censoring $A_{(n)}$' or rc-$A_{(n)}$, which is suitable for right-censored data. In comparison to $A_{(n)}$, rc-$A_{(n)}$ uses the additional assumption that, at the moment of censoring, the residual lifetime of a right-censored unit is exchangeable with the residual lifetimes of all other units that have not yet failed or been censored, see Coolen and Yan [8] for further details of rc-$A_{(n)}$. To formulate the required form of rc-$A_{(n)}$, notation is required for probability mass assigned to intervals without further restrictions on the spread within the intervals. Such a partial specification of a probability distribution is called an M-function [8]. The use of lower and upper probabilities to quantify uncertainty has gained increasing attention during the last decade, short and detailed overviews of theories and applications in reliability, together called 'imprecise reliability', are presented by Coolen and Utkin [6, 21]. Also, Coolen et al. [5] introduced NPI to some reliability applications, including upper and lower survival functions for the next future observation, illustrated with an application with competing risks data. They illustrated the upper and lower marginal survival functions, so each restricted to a single failure mode.

Coolen and Yan [7] presented NPI for comparison of two groups of lifetime data including right-censored observations. Coolen-Maturi et al. [11] extend this for comparing more than two groups in order to select the best group, in terms of largest lifetime. Coolen-Maturi et al. [10] consider selection of subsets of the groups according to several criteria. They allow early termination of the experiment in order to save time and cost, which effectively means that all units in all groups that have not yet failed are right-censored at the time the experiment is ended.

Section 2 of this paper presents a brief overview of NPI for the competing risks problem. NPI for pairwise comparison is introduced in Section 3, presenting the NPI lower and upper probabilities for the event that the lifetime of the next future unit from one group is greater than the lifetime of the next future unit from the second group, with different independent competing risks per group. Our NPI method is illustrated via an example in Section 4. Some concluding remarks are given in Section 5.

§2. NPI for one group with competing risks

In this section, a brief overview of NPI for one group with competing risks is given following the definitions and notations introduced by Maturi et al. [16]. For group X, let us consider the problem of competing risks with J distinct failure modes that can cause a unit to fail. It is assumed that the unit fails due to the first occurrence of a failure mode, and that the unit is withdrawn from further use and observation at that moment. It is further assumed that such

failure observations are obtained for n units, and that the failure mode causing a failure is known with certainty. In the case where the unit did not fail it is right-censored.

Let the failure time of a future unit be denoted by X_{n+1}, and let the corresponding notation for the failure time including indication of the actual failure mode, say failure mode j ($j = 1, \ldots, J$), be $X_{j,n+1}$. As the different failure modes are assumed to occur independently, the competing risk data per failure mode consist of a number of observed failure times for failures caused by the specific failure mode considered, and right-censoring times for failures caused by other failure modes. It should be emphasized that it is not assumed that each unit considered must actually fail, if a unit does not fail then there will be a right-censored observation recorded for this unit for each failure mode, as it is assumed that the unit will then be withdrawn from the study, or the study ends, at some known time. Hence rc-$A_{(n)}$ can be applied per failure mode j, for inference on $X_{j,n+1}$. Let the number of failures caused by failure mode j be u_j, $x_{j,1} < x_{j,2} < \ldots, < x_{j,u_j}$, and let $n - u_j$ be the number of the right-censored observations, $c_{j,1} < c_{j,2} < \ldots < c_{j,n-u_j}$, corresponding to failure mode j. For notational convenience, let $x_{j,0} = 0$ and $x_{j,u_j+1} = \infty$. Suppose further that there are s_{j,i_j} right-censored observations in the interval (x_{j,i_j}, x_{j,i_j+1}), denoted by $c_{j,1}^{i_j} < c_{j,2}^{i_j} < \ldots < c_{j,s_{j,i_j}}^{i_j}$, so $\sum_{i_j=0}^{u_j} s_{j,i_j} = n - u_j$. The random quantity representing the failure time of the next unit, with all J failure modes considered, is $X_{n+1} = \min\limits_{1 \le j \le J} X_{j,n+1}$. The NPI M-functions for $X_{j,n+1}$ ($j = 1, \ldots, J$) are [16]

$$M^j(t_{j,i_j^*}^{i_j}, x_{j,i_j+1}) = M_{X_{j,n_j+1}}(t_{j,i_j^*}^{i_j}, x_{j,i_j+1}) = \frac{1}{(n+1)} (\tilde{n}_{t_{j,i_j^*}^{i_j}})^{\delta_{i_j^*}^{i_j}-1} \prod_{\{r: c_{j,r} < t_{j,i_j^*}^{i_j}\}} \frac{\tilde{n}_{c_{j,r}} + 1}{\tilde{n}_{c_{j,r}}} \quad (1)$$

where $i_j = 0, 1, \ldots, u_j$, $i_j^* = 0, 1, \ldots, s_{j,i_j}$ and

$$\delta_{i_j^*}^{i_j} = \begin{cases} 1 & \text{if } i_j^* = 0 \qquad \text{i.e. } t_{j,0}^{i_j} = x_{j,i_j} \quad \text{(failure time or time 0)} \\ 0 & \text{if } i_j^* = 1, \ldots, s_{j,i_j} \quad \text{i.e. } t_{j,i_j^*}^{i_j} = c_{j,i_j^*}^{i_j} \quad \text{(censoring time)} \end{cases}$$

where $\tilde{n}_{c_{j,r}}$ and $\tilde{n}_{t_{j,i_j^*}^{i_j}}$ are the numbers of units in the risk set just prior to times $c_{j,r}$ and $t_{j,i_j^*}^{i_j}$, respectively. The corresponding NPI probabilities are

$$P^j(x_{j,i_j}, x_{j,i_j+1}) = P(X_{j,n_j+1} \in (x_{j,i_j}, x_{j,i_j+1})) = \frac{1}{n+1} \prod_{\{r: c_{j,r} < x_{j,i_j+1}\}} \frac{\tilde{n}_{c_{j,r}} + 1}{\tilde{n}_{c_{j,r}}} \quad (2)$$

where x_{j,i_j} and x_{j,i_j+1} are two consecutive observed failure times caused by failure mode j (and $x_{j,0} = 0$, $x_{j,u_j+1} = \infty$). Maturi et al. [16] considered the lower and upper probabilities for the event that the next unit fails due to a specific failure mode, say mode j^*, that is for the event $X_{j^*,n+1} < \min\limits_{\substack{1 \le j \le J \\ j \ne j^*}} X_{j,n+1}$, for each $j^* = 1, \ldots, J$.

In addition to notation introduced above, let $t_{j,s_{j,i_j}+1}^{i_j} = t_{j,0}^{i_j+1} = x_{j,i_j+1}$ for $i_j = 0, 1, \ldots, u_j - 1$. For a given failure mode j ($j = 1, \ldots, J$), the NPI lower survival func-

tion [16] is, for $t \in [t_{j,a_j}^{i_j}, t_{j,a_j+1}^{i_j})$ with $i_j = 0, 1, \ldots, u_j$ and $a_j = 0, 1, \ldots, s_{j,i_j}$,

$$\underline{S}_{X_{j,n+1}}(t) = \frac{1}{n+1}\, \tilde{n}_{t_{j,a_j}^{i_j}} \prod_{\{r:c_{j,r}<t_{j,a_j}^{i_j}\}} \frac{\tilde{n}_{c_{j,r}}+1}{\tilde{n}_{c_{j,r}}} \tag{3}$$

and the corresponding NPI upper survival function [16] is, for $t \in [x_{i_j}, x_{i_j+1})$ with $i_j = 0, 1, \ldots, u_j$,

$$\overline{S}_{X_{j,n+1}}(t) = \frac{1}{n+1}\, \tilde{n}_{x_{j,i_j}} \prod_{\{r:c_{j,r}<x_{j,i_j}\}} \frac{\tilde{n}_{c_{j,r}}+1}{\tilde{n}_{c_{j,r}}} \tag{4}$$

Then the lower and upper survival functions for X_{n+1} are given by

$$\underline{S}_{X_{n+1}}^{JCR}(t) = \prod_{j=1}^{J} \underline{S}_{X_{j,n+1}}(t) \quad \text{and} \quad \overline{S}_{X_{n+1}}^{JCR}(t) = \prod_{j=1}^{J} \overline{S}_{X_{j,n+1}}(t) \tag{5}$$

In fact there is a relationship between the above upper survival function in (5) and the upper survival function when all the different failure modes are ignored, that is $\overline{S}_{X_{n+1}}^{JCR}(t) = \overline{S}_{X_{n+1}}(t)$, for more details we refer to Maturi et al. [16].

§3. Pairwise comparison with competing risks

Let X and Y be two independent groups (e.g. treatments) with competing risks $j = 1, \ldots, J$ and $l = 1, \ldots, L$, respectively. These competing risks could be the same (e.g. the lung cancer may affect both men and women independently) or different across the two groups. These competing risks could be observed or unobserved but known, in the sense of not yet having caused any failures (see [9]). For group Y the same notations and definitions as in Section 2 are used, replacing x, u_j, n, c, s, t, i_j, i_j^* by y, v_l, m, d, e, g, i_l, i_l^*, respectively.

In this paper, the main event of interest is that the lifetime of a future unit from group Y is greater than the lifetime of a future unit from group X, i.e. $Y_{m+1} > X_{n+1}$, with J and L independent competing risks affecting group X and group Y, respectively. The following notation is used for the NPI lower and upper probabilities for the event of interest, respectively,

$$\underline{P} = \underline{P}\left(Y_{m+1} > X_{n+1}\right) = \underline{P}\left(\min_{1\le l\le L} Y_{l,m+1} > \min_{1\le j\le J} X_{j,n+1}\right)$$
$$\overline{P} = \overline{P}\left(Y_{m+1} > X_{n+1}\right) = \overline{P}\left(\min_{1\le l\le L} Y_{l,m+1} > \min_{1\le j\le J} X_{j,n+1}\right)$$

These NPI lower and upper probabilities for the event $Y_{m+1} > X_{n+1}$ are

$$\underline{P} = \sum_{C(j,\, i_j)} \left[\prod_{l=1}^{L} \sum_{i_l=0}^{v_l} \sum_{i_l^*=0}^{e_{l,i_l}} 1(g_{l,i_l^*}^{i_l} > \min_{1 \le j \le J} \{x_{j,i_j+1}\}) M^l(g_{l,i_l^*}^{i_l}, y_{l,i_l+1}) \right] \prod_{j=1}^{J} P^j(x_{j,i_j}, x_{j,i_j+1}) \tag{6}$$

$$\overline{P} = \sum_{C(j,\, i_j,\, i_j^*)} \left[\prod_{l=1}^{L} \sum_{i_l=0}^{v_l} 1(y_{l,i_l+1} > \min_{1 \le j \le J} \{t_{j,i_j^*}^{i_j}\}) P^l(y_{l,i_l}, y_{l,i_l+1}) \right] \prod_{j=1}^{J} M^j(t_{j,i_j^*}^{i_j}, x_{j,i_j+1}) \tag{7}$$

where $\sum_{C(j,\, i_j)}$ denotes the sums over all i_j from 0 to u_j for $j = 1, \ldots, J$, and $\sum_{C(j,\, i_j,\, i_j^*)}$ denotes the sums over all i_j^* from 0 to s_{j,i_j} and over all i_j from 0 to u_j for $j = 1, \ldots, J$. And $1(A)$ is the indicator function that equals 1 if A is true and 0 else. The derivation of these NPI lower and upper probabilities will be presented elsewhere.

As mentioned not all these J and L competing risks need to have caused observed failures. Coolen-Maturi and Coolen [9] presented NPI for the case of unobserved failure modes for inference on a single group. Basically, all units, for which data are available, are censored with respect to this unobserved failure mode, and then the corresponding M-functions, introduced in Section 2, are applied per group in order to calculate the NPI lower and upper probabilities from (6) and (7). This will be illustrated via an example in Section 4.

In order to make a decision using our NPI method, we can say that there is strong evidence that the lifetime of a future unit from group Y is likely to be greater than the lifetime of a future unit from group X if $\underline{P}\,(Y_{m+1} > X_{n+1}) > \overline{P}\,(Y_{m+1} < X_{n+1})$, where from the conjugacy property [1] $\overline{P}\,(Y_{m+1} < X_{n+1}) = 1 - \underline{P}\,(Y_{m+1} > X_{n+1})$, and that there is weak evidence for this if $\underline{P}\,(Y_{m+1} > X_{n+1}) > \underline{P}\,(Y_{m+1} < X_{n+1})$ and $\overline{P}\,(Y_{m+1} > X_{n+1}) > \overline{P}\,(Y_{m+1} < X_{n+1})$.

We can also compare the two groups with competing risks using the lower and upper survival functions, (5), namely $\underline{S}_{X_{n+1}}^{JCR}$, $\overline{S}_{X_{n+1}}^{JCR}$, $\underline{S}_{Y_{m+1}}^{LCR}$ and $\overline{S}_{Y_{m+1}}^{LCR}$. The lower and upper survival functions for the case of unobserved failure modes are presented in Coolen-Maturi and Coolen [9]. This will also be illustrated in Section 4.

§4. Example

The original data, used by Park and Kulasekera [17], consist of failure or censoring times for 139 appliances (36 in Group I, 51 in Group II and 52 in Group III) subject to a lifetime test, where a unit is subject to fail due to one of 18 different modes. To clearly illustrate our NPI method, we will use part of this dataset, namely for appliances with lifetimes less than 250. The reduced dataset, in Table 1, consists of 26 appliances (8 in Group I, 11 in Group II and 7 in Group III) where failure mode 11 (FM11) appears at least once across the three groups. FM0 indicates a right censoring time. Table 2 gives the NPI lower and upper probabilities for two cases of interest: In case A, we compare the groups by taking into account all the observed failure modes, while in case B we compare the groups such that, for each group, we

Group I		Group II		Group III	
Time	FM	Time	FM	Time	FM
12	13	45	1	90	1
16	10	47	11	90	11
16	12	73	11	90	11
46	3	136	6	190	1
46	6	136	0	218	0
52	6	136	0	218	0
98	6	136	0	241	1
98	11	136	0		
		145	11		
		190	0		
		190	0		

Table 1: Appliances with lifetimes less than 250

Case	$\underline{P}, \overline{P}$ (II>I)	$\underline{P}, \overline{P}$ (III>I)	$\underline{P}, \overline{P}$ (III>II)
A	(0.5944, 0.9724)	(0.5993, 0.9890)	(0.2914, 0.7840)
B	(0.7222, 0.9074)	(0.6944, 0.9167)	(0.3437, 0.7083)

Table 2: The NPI lower and upper probabilities

re-grouped all observed failure modes in one failure mode. So in case B we have one failure mode per group, which is coincided with the results obtained by Coolen and Yan [7] and it is a special case of the results presented by Coolen-Maturi et al. [11].

We can notice for all events (i.e. II>I, III>I and III>II) that the lower (upper) probabilities for case A are less (greater) than the lower (upper) probabilities for case B. So the imprecision in case B is smaller than that for case A. That is study the data in more details (larger number of competing risks) leads to more imprecision. Increased imprecision if data are included in more details in the NPI approach is a topic of foundational interest that has been observed and discussed before, see Coolen and Augustin [4] and Maturi et al. [16].

From Table 2, we can say that we have strong evidence that the lifetime of a future unit from group I is less than the lifetime of a future unit from group II and III for both cases. On the other hand, for both cases we have weak evidence that the lifetime of a future unit from group II is less than the lifetime of a future unit from group III.

We can also compare these groups using the lower and upper survival functions, see Figure 1. In Figure 1, we provide the NPI lower and upper survival functions for case A where the first graph represents the lower and upper survival functions for the next units from groups I and II, the second graph represents the lower and upper survival functions for the next units from groups I and III and the third graph represents the lower and upper survival functions for the next units from groups II and III. Figure 1 shows indeed that the lifetime of a future unit from group I is likely to be less than the lifetime of a future unit from group II and III. However, we have weak evidence that the lifetime of a future unit from group II is less than the lifetime of a future unit from group III, and we see that the lower (and upper) survival functions for these groups cross each other.

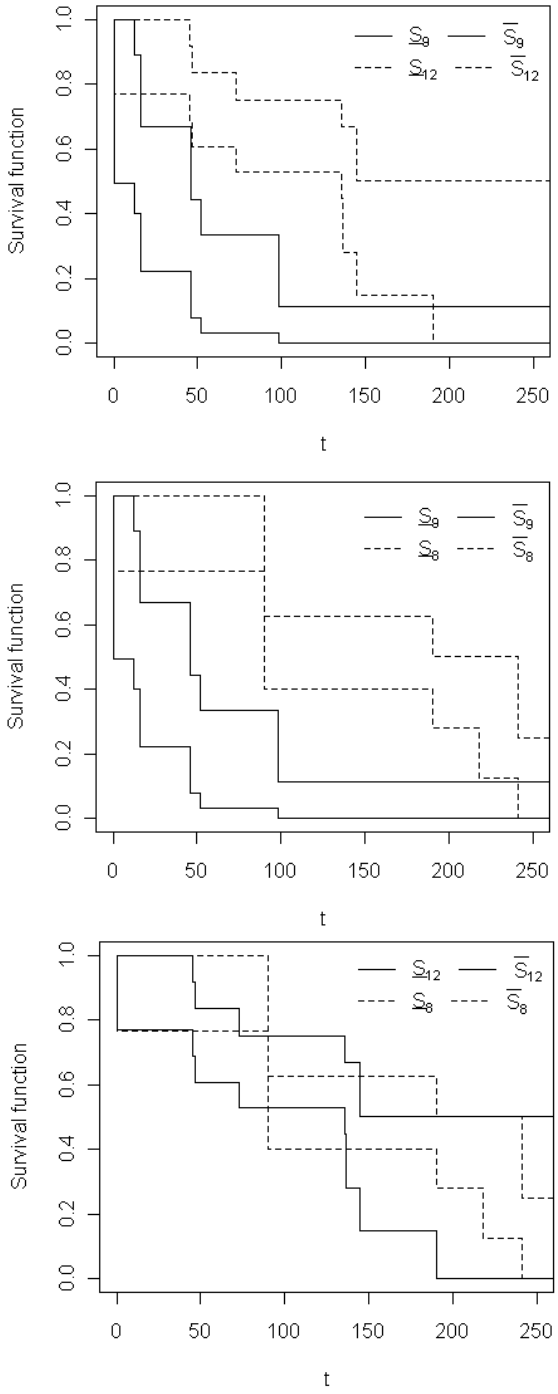


Figure 1: The lower and upper survival functions for case A

§5. Concluding remarks

In this paper we presented NPI for pairwise comparison where each group is subject to several competing risks. We introduced NPI lower and upper probabilities for the event that the lifetime of the next unit from one group is greater than the lifetime of the next unit from the second group, taking into account these competing risks. We found that studying the data in details, so more competing risks, will lead to more imprecision.

References

[1] AUGUSTIN, T., COOLEN, F. P. A. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference 124*, 2 (2004) 251–272.

[2] BEDFORD, T., ALKALI, B., BURNHAM, R. Competing risks in reliability. In: Melnick, E. L., Everitt, B. S. (Eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment*. Chichester: Wiley, 2008, pp. 307–312.

[3] COOLEN, F. P. A. On nonparametric predictive inference and objective bayesianism. *Journal of Logic, Language and Information 15*, 1-2 (2006), 21–47.

[4] COOLEN, F. P. A., AUGUSTIN, T. A nonparametric predictive alternative to the imprecise dirichlet model: the case of a known number of categories. *International Journal of Approximate Reasoning 50*, 2 (2009), 217–230.

[5] COOLEN, F. P. A., COOLEN-SCHRIJNER, P., YAN, K. J. Nonparametric predictive inference in reliability. *Reliability Engineering & System Safety 78*, 2 (2002), 185–193.

[6] COOLEN, F. P. A., UTKIN, L. V. Imprecise reliability. In: Everitt, B. S., Melnick, E. L. (Eds.), *Encyclopedia of Quantitative Risk Analysis and Assessment. Chichester*, Wiley, 2008, pp. 875–881.

[7] COOLEN, F. P. A., YAN, K. J. Nonparametric predictive comparison of two groups of lifetime data. In: *ISIPTA'03: Proceedings of the Third International Symposium on Imprecise Probabilities and their Applications*, Bernard, J. M., Seidenfeld, T., Zaffalon, M. (Eds.), 2003, pp. 148–161.

[8] COOLEN, F. P. A., YAN, K. J. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference 126*, 1 (2004), 25–54.

[9] COOLEN-MATURI, T., COOLEN, F. P. A. Unobserved, re-defined, unknown or removed failure modes in competing risks. *Journal of Risk and Reliability* 225, 4 (2011), 461–474.

[10] COOLEN-MATURI, T., COOLEN-SCHRIJNER, P., COOLEN, F. P. A. Nonparametric predictive selection with early experiment termination. *Journal of Statistical Planning and Inference 141*, 4 (2011), 1403–1421.

[11] COOLEN-MATURI, T., COOLEN-SCHRIJNER, P., COOLEN, F. P. A. Nonparametric predictive multiple comparisons of lifetime data. *Communications in Statistics - Theory and Methods 41*, 22 (2012), 4164–4181.

[12] DE FINETTI, B. Theory of Probability: A Critical Introductory Treatment. Wiley, London, 1974.

[13] HILL, B. M. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association 63*, 322 (1968), 677–691.

[14] HILL, B. M. De finetti's theorem, induction, and a_n, or bayesian nonparametric predictive inference (with discussion). In: *Bayesian Statistics 3*. Bernando, J. M., DeGroot, M. H., Lindley, D. V., Smith, A. (Eds.) Oxford University Press, 1988, pp. 211–241.

[15] LUO, X., TURNBULL, B. W. Comparing two treatments with multiple competing risks endpoints. *Statistica Sinica 9*, 4 (1999), 985–997.

[16] MATURI, T. A., COOLEN-SCHRIJNER, P., COOLEN, F. P. A. . Nonparametric predictive inference for competing risks. *Journal of Risk and Reliability 224*, 1 (2010b), 11–26.

[17] PARK, C., KULASEKERA, K. B. Parametric inference of incomplete data with competing risks among several groups. *IEEE Transactions on Reliability 53*, 1 (2004), 11–21.

[18] RAY, M. R. Competing risks. In: Melnick, E. L., Everitt, B. S. (Eds.), Encyclopedia of Quantitative Risk Analysis and Assessment. Chichester: Wiley, 2008, pp. 301–307.

[19] SARHAN, A. M., HAMILTON, D. C., SMITH, B. Statistical analysis of competing risks models. *Reliability Engineering & System Safety 95*, 9 (2010), 953–962.

[20] TSIATIS, A., A nonidentifiability aspect of the problem of competing risks. In: *Proceedings of the National Academy of Sciences of the United States of America 72*, 1975, pp. 20–22.

[21] UTKIN, L. V., COOLEN, F. P. A. Imprecise reliability: An introductory overview. In: *Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability*. Levitin, G. (Ed.), Springer, New York, 2007, pp. 261–306.

[22] WALLEY, P. Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, London, 1991.

[23] WEICHSELBERGER, K. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning 24*, 2-3 (2000) 149–170.

[24] WEICHSELBERGER, K. Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept. *Physika*, Heidelberg, 2001.

Tahani Coolen-Maturi
Durham University Business School
Durham University
Durham, DH1 3LB, UK
tahani.maturi@durham.ac.uk

Monografías Matemáticas García de Galdeano **38**, 55–64 (2013)

A Comparative Study of Bullwhip Effect in a Multi-Echelon Forward-Reverse Supply Chain

Debabrata Das and Pankaj Dutta

Abstract. Along with the forward supply chain organization needs to consider the impact of reverse logistics due to social awareness, environmental benefits and economic advantages. An important observation in a supply chain management, known as bullwhip effect, refers to the phenomenon where orders to the supplier tend to have larger variance than sales to the buyer (e.g., demand distortion), and the distortion propagates upstream in an amplified form (e.g., variance amplification) [1]. The quality and quantity of used products return to the collection points are uncertain in the reverse channel. Because of this, the systematic distortion is inevitable and bullwhip effect may occur at retailer, distributor and manufacturer level. In this paper; first, we propose a system dynamics framework for a multi-echelon integrated forward-reverse supply chain. Then, in the simulation study, we analyze the order variation at both retailer and distributor level and compare the bullwhip effects of an integrated forward-reverse supply chain with that of a traditional forward supply chain. Also, in the proposed model, a sensitivity analysis is performed to examine the impact of inventory adjustment time and inventory cover time on the order variance and bullwhip effect.

Keywords: Reverse Supply Chain, Bullwhip Effect, Simulation, System Dynamics, Return Rate, Inventory Cover Time, Inventory Adjustment Time

AMS classification: 37M05, 90C31, 90B50

§1. Introduction

A large number of successful companies focuses on forward supply chain but experience a lack of control over their reverse logistics process which leads to higher cost, poor customer service, reduced asset recovery and most importantly, environmental disaster. Along with the forward supply chain organization needs to consider the impact of reverse logistics due to the potentials of value recovery from the used products, social awareness and strict legislations especially for electronics and automobiles industries [2]. Pagell et al. [3] pointed out that product remanufacturing is the most desirable option for end-of-life product management than a scrap or spares recovery since it minimizes the environmental impacts, results in lower loss of value, and can create new market opportunities. Fleischmann et al. [4] provided a review of the quantitative models for reverse logistics in which they reported that most of the papers in the area of integrated reverse logistics are confined to single issues such as network design, shop-floor control and inventory control, while comprehensive approaches are rare

as variety of factors are involved in a general framework and the complexity of their interdependencies. Furthermore, long-term strategic management problems of integrated reverse logistics systems have not been studied extensively. System dynamics (SD) is a powerful methodology for obtaining the insights of these kinds of problems having dynamic complexity; but there are very few literatures which modeled the integrated aspects of forward and reverse supply chain using SD. Spengler and Schroter [5] modeled an integrated production and recovery system for supplying spare parts using SD to evaluate various strategies. Georgiadis and Vlachos [6] developed a SD model to evaluate the effect of environmental issues on long-term decision making in collection and remanufacturing activities.

As the quality and quantity of used products return to the collection points are uncertain in the reverse channel, the systematic distortion is inevitable and bullwhip effect may occur at retailer, distributor and manufacturer level [7]. Almost all quantitative literature is based upon traditional supply chain. There are only few papers studied the order variations and bullwhip effects in an integrated RL framework. Zhou and Disney [8] studied the bullwhip effects and order variations in a closed-loop supply chain based on the control theory. Pati et al. [9] developed an analytical expression for measuring the bullwhip effect in a six echelon closed-loop supply chain for recycling products like paper, plastic, etc. In this research work, we simulate the order variation of different logistics participants over time and compare the bullwhip effects of the traditional forward supply chain with that of integrated forward-reverse supply chain. Also, sensitivity analysis is performed to examine the impact of inventory adjustment time and inventory cover time on the order variance and bullwhip effect.

§2. Integrated Forward-Reverse Supply Chain

The process of reverse supply chain is more complicated than forward supply chain since return flows may include several activities such as collection, checking, sorting, disassembly, remanufacturing, disposal and redistribution [4]. Moreover, the quality and quantity of used products return to the collection points are uncertain in the reverse channel. In this research work, we focus on a single product integrated forward-reverse supply chain (see Fig. 1).

2.1. System Dynamics Model

Forrester [10] introduced SD as a modelling and simulation methodology for framing, understanding, and discussing complex issues and problems. The SD methodology is described by causal-loop diagrams. The structure of a SD model contains stock (state), flow (rate) and auxiliary/constant variables. Stock variables are the accumulations (e.g. inventories) within the system. The flow variables represent the flows in the system (e.g. remanufacturing rate) from one stock to another. With a causal loop diagram, the stock and flow diagram shows relationships among variables which have the potential to change over time. The mathematical formulation consists of a system of differential equations, which is numerically solved via simulation. Nowadays, high-level graphical simulation programs support the analysis and study of these systems. These programs include Vensim, i-think and Powersim etc. Here, we choose Vensim (version: windows 5.10 e) as a tool to simulate the model.

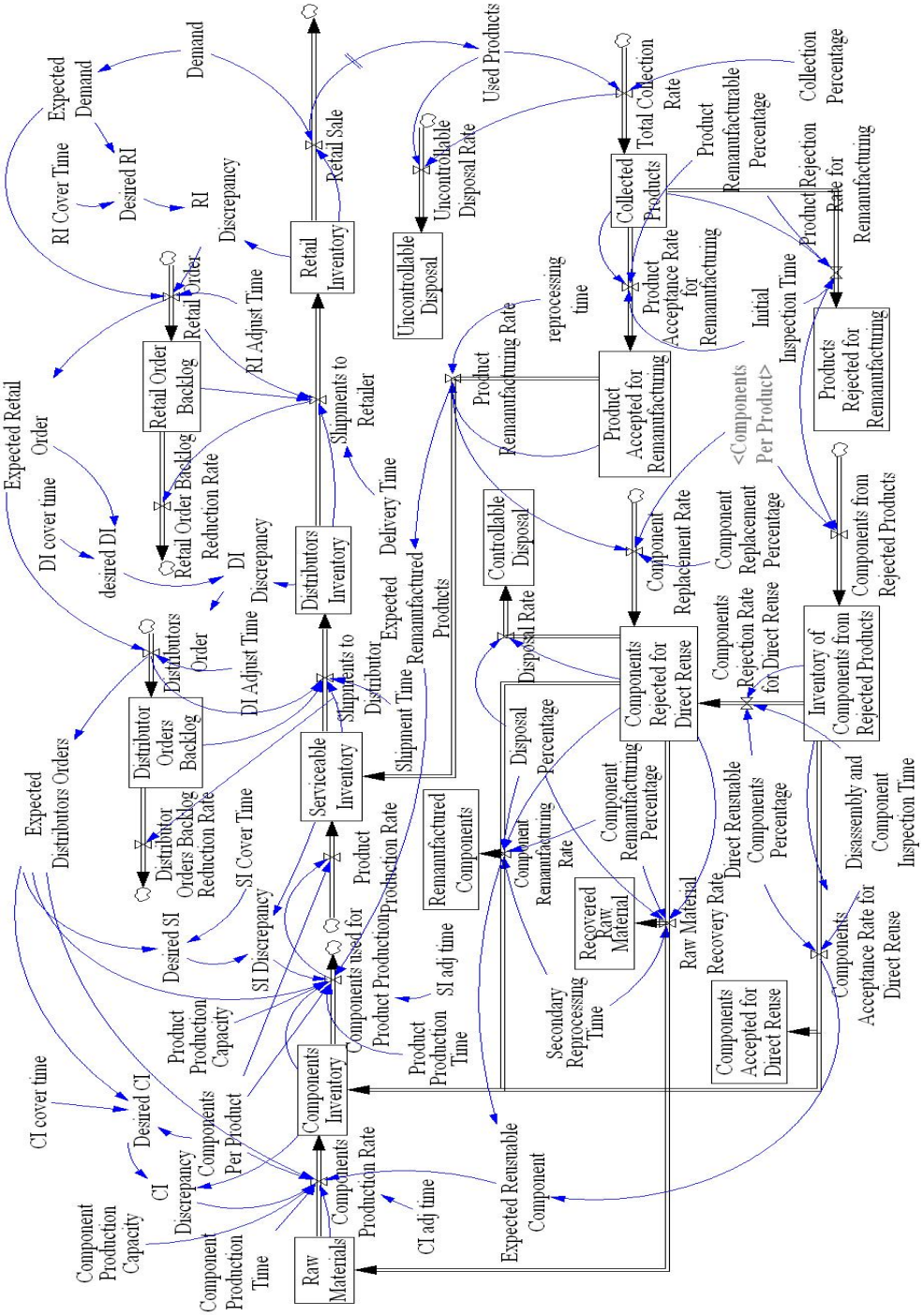


Figure 1: Stock-flow diagram of the integrated forward-reverse supply chain

2.1.1. Stock-Flow Diagram

The first step of the analysis is to capture the relationships among the system operations in a SD manner. Fig. 1 depicts the stock-flow diagram of the integrated forward-reverse supply chain in which all the stock variables are represented by "*box*" symbol and flow variables by "*hour glass*" symbol. Throughout the paper, variables are described in *Italic* font. Important equations related to production, inventory, transportation, order, collection and remanufacturing process are presented in the Appendix II. Because of the high complexity of the integrated forward-reverse supply chain, it can be divided into the following subsystems:

Forward Supply Chain: The forward supply chain begins from the upper left corner of Fig. 1 and it comprises three echelons: producer, distributor and retailer. Specifically, the new products are first transferred from the producer to the distributor then to the retailer and finally sold to the customer to satisfy the demands. The producer's demand for raw materials (*Raw Materials*) is satisfied with a mix of fresh raw materials, and recovered raw materials (*Raw Material Recovery Rate*) deriving from the firm's recycling operations. *Components Production Rate* depletes raw materials and increase *Components Inventory* (CI). *Product production rate* depletes *Components Inventory* and increase *Serviceable Inventory* (SI). *Shipments to Distributor* deplete *Serviceable Inventory* and increase *Distributors Inventory* (DI). In the same way, products delivered from the upper stream increase the inventory of retailer (*Retail Inventory*), which can satisfy the *Demand* of end-users. One important term of *Components (or Product) Production Rate* is *CI (or SI) Adj time* that represents how quickly the firm tries to correct the discrepancy in component (or product) inventory.

The remanufacturing process supplements the production process of the forward channel. Producer's requirement for components is satisfied with a mix of new components produced by firm, and *Remanufactured Components* derived from remanufacturing process. Similarly, producer's requirement for products is satisfied with a mix of new products produced by firm, and *Remanufactured Products* derived from *Collected Products*. In this paper, we assume that demand of end-user is lost if it is not satisfied in the current period. Although, *Distributor Orders Backlog* and *Retailer Orders Backlog* are satisfied in a future period.

Reverse Supply Chain: In the reverse channel, we address the recovery process in three distinct ways, namely; product remanufacturing, component reuse and remanufacturing, and raw material recovery. We assume that remanufacturing activity can bring the products and components back into an "as good as new" condition by carrying out the necessary disassembly, overhaul and replacement operations. In the simulation study, it is assumed that there is no constraint on the capacity of collection, inspection, sorting and restoring.

Sold products after their uses turn into used products. Then, *Used Products* are either uncontrollably disposed (*Uncontrollable Disposal*) or collected for reuse (*Collected Products*). After the initial inspection, if the *Collected Products* are accepted for remanufacturing (*Product Accepted for Remanufacturing*), then with some reprocessing, the remanufactured products (*Product Remanufacturing Rate*) can be added to the serviceable inventory of forward channel. If the products are not in a condition to remanufacture, then it is disassembled into various components. During the process of product remanufacturing, if new replacement (*Component Replacement Rate*) is required for some components, then the old components

Variance ($unit^2$)	Integrated Supply Chain	Traditional Supply Chain
Actual demand	9326	9326
Order at retailer	17111	31940
Order at distributor	38264	80558

Table 1: Comparison of order variance: integrated vs traditional supply chain

	Retailer Level		Distributor Level	
	Integrated SC	Traditional SC	Integrated SC	Traditional SC
Bullwhip Effect	1.83	4.10	3.42	8.64

Table 2: Comparison of bullwhip effects in an integrated and traditional supply chain

are processed further for component remanufacturing and/or raw material recovery.

In the model, it is assumed that the disassembled components can have three categories: one is direct reusable components (*Components Accepted for Direct Reuse*) that can be directly used to increase the *Components Inventory* in the forward channel; the second is the part of *Components Rejected for Direct Reuse* which requires further reprocessing. After reprocessing, the *Remanufactured Components* can be used to increase the *Components Inventory* in the forward channel. The third is rejected components that does not survive the first two screening levels but can be used either for raw material recovery (*Recovered Raw Material*) to increase the *Raw Materials inventory* in the forward channel or sent directly for *Controllable Disposal*.

§3. Results and Discussions

In this section, we demonstrate the behavior analysis of the integrated forward-reverse supply chain and discuss some of the important results. Although, we develop a generalized framework but to analyze the performance of the proposed integrated system, values of the most of the parameters are chosen as in [11]. Some of the important parameters associated with the SD model are presented in the Appendix I. The length of the time horizon is 300 weeks for the simulation.

3.1. Bullwhip Effects and Order Variations

We use the corresponding stock-flow diagram (see Fig. 1) of the integrated forward-reverse supply chain to simulate its system performance and compare the bullwhip effect of the integrated forward-reverse supply chain with that of traditional (i.e. only forward) supply chain. Table 1 shows the variance of actual demand and variance of order placed by retailer and distributor in an integrated forward-reverse and traditional supply chain. It is very clear that variance of orders at both retailer and distributor is much higher in a traditional supply chain compared to that of in an integrated forward-reverse supply chain.

We compute the bullwhip effect of the systems using the following formulation given by [12]: Bullwhip Effect = Var (Order Rate) / Var (Demand) and make a comparison of

Cover/Adjustment time (week)		1	2	3	4
Bullwhip Effect due to RICT	Retailer	1.16	3.03	5.62	10.75
	Distributor	1.94	7.67	15.09	30.30
Bullwhip Effect due to RIAT	Retailer	3.56	1.83	1.46	1.31
	Distributor	9.40	4.10	2.89	2.40
Bullwhip Effect due to DICT	Retailer	1.84	1.65	1.22	1.19
	Distributor	2.11	6.24	9.12	18.30
Bullwhip Effect due to DIAT	Retailer	1.83	1.83	1.83	1.83
	Distributor	7.54	4.10	3.25	2.85

Table 3: Sensitivity analysis of bullwhip effect at retailer and distributor

bullwhip effect at both retailer and distributor level for an integrated forward-reverse supply chain (SC) and traditional supply chain (SC) in Table 2. It is clear that the bullwhip effect at both retailer and distributor level in the traditional supply chain is much higher than that of integrated forward-reverse supply chain. Important reason behind this phenomenon is that the remanufactured products, reusable components and remanufactured components derived from firm's recycling process supplement the product inventory and component inventory of producer in the forward channel which helps in reducing the inventory discrepancy at various stages of the forward supply chain. So, the results indicate that the remanufacturing process reduces the bullwhip effects of different logistics participants in a supply chain.

3.2. Sensitivity Analysis

In this section, we investigate the impact of various system parameters on the performance of the proposed integrated forward-reverse supply chain model. Conducting a detailed sensitivity analysis by taking into account all the system parameters is hardly possible. Hence, in this study, we concentrate on examining the impact of inventory cover time and inventory adjustment time on bullwhip effect at both retailer and distributor level. Inventory cover time describes a level of extra stock that is maintained to mitigate risk of stock outs due to uncertainties in supply and demand. Inventory adjustment time represents how quickly a firm tries to correct the discrepancy between desired serviceable inventory and actual serviceable inventory. The result in Table 3 shows that the bullwhip effect increases at both retailer and distributor level as the retail inventory cover time (RICT) increases. On the other hand, bullwhip effect decreases at both retailer and distributor levels as the retail inventory adjustment time (RIAT) increases which is due to the fact that if a firm adjusts the discrepancy between desired serviceable inventory and actual serviceable inventory very quickly, then the variations in order increases. From the last two rows of Table 3, it can be seen that the bullwhip effect increases at distributor level with the increment of distributor inventory cover time (DICT) and the bullwhip effect decreases with the increment of distributor inventory adjustment time (DIAT); but the changes of cover time and adjustment time in distributor level has almost no impact in determining the bullwhip effect at retailer level.

§4. Conclusion

In this paper, we have proposed a SD framework for a multi-echelon forward-reverse supply chain. We analyzed the order variation at both retailer and distributor level and compare the bullwhip effects of different logistics participants over time between the traditional forward supply chain and the integrated forward-reverse supply chain. Our results showed that the inclusion of remanufacturing can reduce the order variation and bullwhip effect in an integrated forward-reverse system. Also, sensitivity analysis is performed to examine the impact of inventory adjustment time and inventory cover time on the order variance and bullwhip effect. The developed model can be used to conduct various "what-if" analysis thus identifying efficient policies and further to answer questions about the long-term operation of the integrated forward-reverse supply chains. The proposed SD framework can be extended by including the associated costs which helps to measure the economic performance of the integrated supply chain. Additionally, the uncertainty issues associated with the collection of used products can be addressed as a future work.

§A. Appendix I - Model Parameters

CI Cover Time = 1.5 week
CI Adj Time = 2 week
SI Cover Time = 1.5 week
SI Adj Time = 2 week
DI Cover Time = 1.5 week
DI Adjust Time = 2 week
RI Cover Time = 1.5 week
RI Adjust Time = 2 week
Component Production Time = 1.2 week
Product Production Time = 2 weeks
Shipment Time = 1.5 week
Delivery Time = 1.5 week
Initial Inspection Time = 1 week
Reprocessing Time (Product) = 1.2 week
Secondary Reprocessing Time = 2 week
Disassembly and Component Inspection Time = 1 week
Components Per Product = 3
Product Production Capacity = 700
Component Production Capacity = 2100
Cycle Life of Product = 50 weeks
Demand = Random Normal (650, 100)
Collection Percentage = 50%
Product Remanufacturable Percentage = 80%
Component Remanufacturing Percentage = 70%
Component Replacement Percentage = 15 %
Direct Reusable Component Percentage = 65%
Disposal Percentage = 10%

§B. Appendix II - Model Equations

B.1.

The important equations related to component and product production rate are following:

*Components Production Rate = MAX (MIN (MIN (Raw Materials / Component Production Time, (Expected Distributors Orders*Components Per Product - Expected Reusable Component + CI discrepancy / CI Adj Time)), Component Production Capacity), 0)*

Expected reusable components = SMOOTH (Component Remanufacturing Rate + Components Acceptance Rate for Direct Reuse, 1)

CI discrepancy= MAX (Desired CI - Components Inventory, 0)

Components Inventory = INTEGRATION (Components Production Rate + Component Remanufacturing Rate + Components Acceptance Rate for Direct Reuse) - Components used for Product Production)

Product Production Rate = Components used for Product Production / Components Per Product

*Components used for Product Production = MAX (MIN (MIN (Components Inventory / Product Production Time, Product Production Capacity*Components Per Product), (Expected Distributors Orders - Expected Remanufactured Products + SI discrepancy / SI Adj Time)*Components Per Product) , 0)*

Expected Remanufactured Products = SMOOTH (Product Remanufacturing Rate, 1)

SI discrepancy= MAX (Desired SI - Serviceable Inventory, 0)

Serviceable Inventory = INTEGRATION (Product Production Rate + Product Remanufacturing Rate - Shipments to Distributor)

B.2.

The important equations related to inventory, transportation and order are following:

Distributors Inventory = INTEGRATION (Shipments to Distributor - Shipments to Retailer)

Shipments to Distributor = IF THEN ELSE (Serviceable Inventory - Distributors Order - Distributor Orders Backlog >=0, Distributors Order + Distributor Orders Backlog, Serviceable Inventory) / Shipment Time

Distributor Orders Backlog = INTEGRATION (Distributors Order - Distributor Orders Backlog Reduction Rate)

Distributors Order = Expected Retail Order + DI discrepancy / DI Adjust Time

Distributor Orders Backlog Reduction Rate = Shipments to Distributor

B.3.

The important equations related to collection and remanufacturing are following:

*Total Collection Rate = Used Products*Collection Percentage*

Product Accepted for Remanufacturing = INTEGRATION (Product Acceptance Rate for Remanufacturing - Product Remanufacturing Rate)

Product Remanufacturing Rate = Product Accepted for Remanufacturing / Reprocessing Time

Components Acceptance Rate for Direct Reuse = Inventory of Components from Rejected Products Direct Reusable Components Percentage / Disassembly and Component Inspection Time*

Component Remanufacturing Rate = (Components Rejected for Direct Reuse)(1-Disposal Percentage) *Component Remanufacturing Percentage / Secondary Reprocessing Time*

Recovered Raw Material = INTEGRATION (Raw Material Recovery Rate)

Raw Material Recovery Rate = Components Rejected for Direct Reuse(1-Disposal Percentage) *(1-Component Remanufacturing Percentage) / Secondary Reprocessing Time*

Controllable Disposal = INTEGRATION (Disposal Rate)

*Disposal Rate = Components Rejected for Direct Reuse*Disposal Percentage*

References

[1] LEE, H., PADMANABHAN, P., AND WHANG, S. The bullwhip effect in supply chains. *Sloan Management Review* 38 (1997), 93–102.

[2] POKHAREL, S., AND MUTHA, A. Perspectives in reverse logistics: A review. *Resources, Conservation and Recycling* (2009), 175–182.

[3] PAGELL, M., WU, Z., AND NAGESH, N. The supply chain implications of recycling. *Business Horizon* 50 (2007), 133–143.

[4] FLEISCHMANN, M., DEKKER, R., VAN DER LAAN, E., VAN NUMEN, J., VAN WASSENHOVE, L., AND RUWAARD, J. Quantitative models for reverse logistics: A review. *European Journal of Operational Research* 103 (1997), 1–17.

[5] SPENGLER, T.,AND SCHROTER, M. Strategic management of spare parts in closed-loop supply chains: a system dynamics approach. *Interfaces* 33(6) (2003), 7–17.

[6] GEORGIADIS, P., AND VLACHOS, D. The effect of environmental parameters on product recovery. *European Journal of Operational Research* 157(2) (2004), 449–464.

[7] QINGLI, D., HAO, S., AND HUI, Z. Simulation of remanufacturing in reverse supply chain based on system dynamics. *IEEE* (2008). 978-1-4244-1672-1.

[8] ZHOU, L., AND DISNEY, S. Bullwhip and inventory variance in a closed loop supply chain. *OR Spectrum* 28 (2006), 127–149.

[9] PATI, R. K., VRAT, P., AND KUMAR, P. Quantifying bullwhip effect in a closed loop supply chain.*OPSEARCH* 47(4) (2010), 231–253.

[10] FORRESTER, J. Industrial Dynamics.*Cambridge, MA: MIT Press* (1961).

[11] VLACHOS, D., GEORGIADIS, P., AND IAKOVOU, E. A system dynamics model for dynamic capacity planning of remanufacturing in closed-loop supply chains. *Computers and Operations Research* 34 (2) (2007), 367–394.

[12] CHEN, F., DREZNER, Z., RYAN, J. K., AND SIMCHI-LEVI, D. Quantifying the Bullwhip Effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science* 46 (3) (2000), 436–443.

Debabrata Das
Research Scholar
SJM School of Management
Indian Institute of Technology, Bombay
Powai, Mumbai-400076
India
debabrata.das@iitb.ac.in

Pankaj Dutta
Assistant Professor
SJM School of Management
Indian Institute of Technology, Bombay
Powai, Mumbai-400076
India
pdutta@iitb.ac.in

Monografías Matemáticas García de Galdeano **38**, 65–74 (2013)

A PROPOSED MARKOV MODEL FOR PREDICTING THE STRUCTURE OF A MULTI-ECHELON EDUCATIONAL SYSTEM IN NIGERIA

Virtue U. Ekhosuehi and Augustine A. Osagiede

Abstract. This paper is concerned with deriving, using logistic and Markov chain theoretic methodologies, a transition model for a stable educational system in Nigeria. The resulting transition model is the neo-stable imbedded Markov model. It is shown using an entropy-based uncertainty metric that the neo-stable imbedded Markov model is preferable to the stable imbedded Markov model in literature for long-term projection using dataset from a university.

Keywords: binomial logistic model, Markov model, multi-echelon educational system, Nigeria.

AMS classification: 60J20, 97B10.

§1. Introduction

This study is aimed at predicting the structure of educational system in Nigeria. The educational process in Nigeria is characterized by wastage arising, *inter alia*, from financial insolvency and distortions such as incessant strike, students' rampage, etc. Such distortions lead to an extension of the academic calendar. As a consequence, equidistant sessions may not be feasible and the discrete-time homogeneous Markov chain may not give substantive meaning (e.g., raising the transition matrix to a fractional index as $\frac{3}{2}$). Since it is not expedient to assume equal length of sessions, we figure out ways to develop appropriate models which are suited for any time instant. We estimate transition probabilities and derive a stationary continuous-time imbedded Markov chain framework for the educational system. The stationary continuous-time imbedded Markov chain is as defined in [1-2]. We consider an expanding multi-echelon educational system with a set of states $\Re = 0 \cup S$, where the notation 0 denotes the state outside the educational system and $S = \{1, 2, \ldots, k\}$ is the set of levels in the system. It is assumed that the states of the system are non-overlapping and the grades are finite and exhaustive. We also assume that flows within the transient states of the educational system are governed by transition probabilities and that they are random variables with a multinomial distribution [3]. The log-likelihood function of the distribution depends on the model parameters and not on the observed values [4]. We estimate the transition probabilities of the non-homogeneous evolution of the educational system by solving the problem:

P1: Maximize

$$\Psi(p_{i1}(t), p_{i2}(t), \ldots, p_{ik}(t)) = \sum_{j=1}^{k} n_{ij}(t) \log p_{ij}(t) \tag{1}$$

subject to:

$$\log[p_{i0}(t)(\sum_{j=1}^{k} p_{ij}(t))^{-1}] = \boldsymbol{x}_i^{'}(t)\beta_i, \tag{2}$$

$$\sum_{j=1}^{k} p_{ij}(t) + p_{i0}(t) = 1, \tag{3}$$

$$p_{ij}(t) \geq 0, p_{i0}(t) \geq 0, i, j \in S, t = 0, 1, 2, \ldots, T. \tag{4}$$

$\Psi(p_{i1}(t), p_{i2}(t), \ldots, p_{ik}(t))$ is the log-likelihood function of the distribution of the transient states. $p_{ij}(t)$ is the probability of students flow from level i to level j in period t. $p_{i0}(t)$ is the probability of students leaving level i in period t. $\boldsymbol{x}_i^{'}(t) = [1, x_{1i}(t), x_{2i}(t), \ldots, x_{(h-1)i}(t), \ldots, x_{(p-1)i}(t)]$ with $x_{(h-1)i}(t), h = 1, \ldots, p$, being an observation corresponding to the $(h-1)$th systems differential variables in level i in session t. $\mathbf{x}^{'}(t) = [\boldsymbol{x}_1^{'}(t), \boldsymbol{x}_2^{'}(t), \ldots, \boldsymbol{x}_k^{'}(t)]^{'}$ is a $k \times p$ matrix of the systems differential variables. $n_{ij}(t)$ is the number of students moving from level i to level j in session t. T is the maximum period for which data are available. In **P1** school fees and other differential variables (such as promotion criteria and environmental factors e.g. land use mix, traffic zone, etc.) are incorporated into the transition model through the binomial logistic wastage rate (2). The discrete-time imbedded transition matrix resulting from the transition probability estimates in **P1** is called the non-homogeneous empirical transition matrix (NHETM) and denoted as $\mathbf{Q}(t)\Big|_{\mathbf{x}'(t)}$. The NHETM is analogous to the block structure of the discrete-time imbedded Markov chain in [5-6]. Osagiede and Ekhosuehi [7] had earlier developed a model for homogeneous Markov systems based on the imbedded Markov chain [5]. We refer to the model of Osagiede and Ekhosuehi [7] as the stable imbedded Markov (SIM) model. The term 'stable' means that the enrolment stock expands deterministically at a constant growth rate [3]. We develop a stationary continuous-time imbedded Markov chain framework to extrapolate the shifts in structure. Since data in an educational system may be obtained at equidistant intervals, it is rational for the transition matrix of our proposed model to be as close as the discrete-time imbedded Markov chain formulation. In this light we solve the quasi-imbedding problem:

P2: Find an intensity matrix $\mathbf{G} = (g_{ij})$ such that: $\prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)} \sim \exp([t+1]\mathbf{G}), t \geq 5$, $g_{ij} \geq 0$ for $i \neq j, \sum_{j=1}^{k} g_{ij} = 0, i \in S$.

The twindle sign $\sim$ means that $\prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)}$ approaches $\exp([t+1]\mathbf{G})$, where $\mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)}$ is the quasi-non-homogeneous imbedded Markov chain (qNHIMC). Problem **P2** is analogous to the one in [2] and $t \geq 5$ is in line with [8]. The definition of problem **P2** circumvents the challenge that would have been posed by the sparse block structure of empirical transition matrices. By solving **P2**, we attempt to bridge the gap between the discrete [5-6] and

the continuous-time imbedded Markov chain formulations [1-2]. Using the solution to **P2**, we develop a transition model which we call the neo-stable imbedded Markov (neo-SIM) model. Throughout the paper, the notations t and ς are used as discrete indices, while τ_v, $v = 1, 2, \cdots$, is used as a continuous index.

§2. Methodology

In this section, we make propositions about the transition probabilities and the expected structure of the system viz-a-viz the solutions to **P1** and **P2**.

Proposition 1: *In a $k-$echelon educational system where the flows satisfy the multinomial distribution and the wastage rates vary in a binomial logistic manner with the differential variables of the school, the transition probabilities are estimated as*:

$$\hat{p}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} = \begin{cases} [1+\exp(\boldsymbol{x}_i'(t)\beta_i^*)]^{-1}\exp(\boldsymbol{x}_i'(t)\beta_i^*), & i \in S, j = 0 \\ \frac{n_{ij}(t)}{\sum_{j=0}^{k} n_{ij}(t)}[1+\exp(\boldsymbol{x}_i'(t)\beta_i^*)]^{-1}, & i, j \in S \end{cases},$$

provided $\det\left[\frac{\partial^2 L}{\partial\beta_i\partial\beta_i'}\right] \neq 0$, *for each* $\beta_i = \beta_i^{(\gamma)}, \gamma = 0, 1, \cdots$, *where* $\hat{p}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)}$ *is the estimated probability of students flow from level i to level j given $\boldsymbol{x}_i'(t)$ for each t and L is the log-likelihood function of the binomial logistic constraint.*

Proof: From the constraints (2) and (3) of **P1**, we obtain

$$\log(\sum_{j=1}^{k} p_{ij}(t)) = -\log(1+\exp(\boldsymbol{x}_i'(t)\beta_i)). \tag{5}$$

Let $z_i^m(t)$ be a binary random variable defined as:

$$z_i^m(t) = \begin{cases} 1 & \text{if a student } m \text{ leaves level } i \text{ in session } t \\ 0 & \text{if student } m \text{ does not leave level } i \text{ in session } t \end{cases},$$

with probabilities $Prob(z_i^m(t) = 1) = p_{i0}(t)$ and $Prob(z_i^m(t) = 0) = \sum_{j=1}^{k} p_{ij}(t)$, for all $i \in S$. Let $w_i(t)$ be a random variable which represents the total number of wastage in level i in session t. If $n_i(t)$ is the total number of students enrolled in level i in session t, then $w_i(t) = \sum_{m=1}^{n_i(t)} z_i^m(t)$. The distribution of the random variable $w_i(t)$ is a binomial distribution [4]. The log-likelihood function of the binomial distribution is

$$L = \sum_{t=1}^{T}\left(n_{i0}(t)\log[p_{i0}(t)(\sum_{j=1}^{k} p_{ij}(t))^{-1}] + n_i(t)\log(\sum_{j=1}^{k} p_{ij}(t)) + \log\begin{pmatrix} n_i(t) \\ n_{i0}(t) \end{pmatrix}\right), \tag{6}$$

where $n_{i0}(t)$ is the wastage flow from level i in session t. From constraint (2) and Eq. (5),

Eq. (6) becomes

$$L = \sum_{t=1}^{T} \left(n_{i0}(t)\boldsymbol{x}_i^{'}(t)\beta_i - n_i(t)\log[1+\exp(\boldsymbol{x}_i^{'}(t)\beta_i)] + \log\left(\begin{array}{c} n_i(t) \\ n_{i0}(t) \end{array} \right) \right). \tag{7}$$

Let $\mathbf{U}(\beta_i) = \left(\frac{\partial L}{\partial \beta_{1i}}, \cdots, \frac{\partial L}{\partial \beta_{pi}} \right)^{'}$. We obtain a solution to the problem $\mathbf{U}(\beta_i) = \mathbf{0}$, which is nonlinear in β_i, employing the iteratively reweighted least squares algorithm [9]. The method involves solving repeatedly

$$\beta_i^{(\gamma+1)} = \beta_i^{(\gamma)} - [\Im(\beta_i^{(\gamma)})]^{-1}\mathbf{U}(\beta_i^{(\gamma)}), \gamma = 0, 1, \cdots, \tag{8}$$

provided $\det[\Im(\beta_i^{(\gamma)})] \neq 0$, where $\Im(\beta_i^{(\gamma)}) = \left[\frac{\partial^2 L}{\partial \beta_i^{(\gamma)} \partial \beta_i^{(\gamma)'}} \right]$ is a $p \times p$ matrix of second-order partial derivatives of the log-likelihood function, L, evaluated at the γth iteration around β_i. The iteration is started from $\beta_i^{(0)} = \mathbf{0}$ and it stops when the parameter estimates do not change significantly any more. Let β_i^* be the numerical solution to the system in Eq. (8). After some simplifications using β_i^* and constraint (2), the objective function (1) is rewritten as

$$\Psi(p_{i1}(t), \cdots, p_{ik}(t)) =$$

$$\sum_{j=1}^{k-1} n_{ij}(t)\log p_{ij}(t) + n_{ik}(t)\log([1+\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*)]^{-1} - \sum_{j=1}^{k-1} p_{ij}(t)). \tag{9}$$

Thus, we obtain, after taking the partial derivatives of $\Psi(p_{i1}(t), \cdots, p_{ik}(t))$ with respect to each $p_{ij}(t)$ and then setting the derivatives to zero,

$$\hat{p}_{i0}(t)\Big|_{\boldsymbol{x}_i'(t)} = [1+\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*)]^{-1}\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*), \tag{10}$$

$$\sum_{j=1}^{k} \hat{p}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} = [1+\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*)]^{-1}, \tag{11}$$

and

$$\hat{p}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} = \frac{n_{ij}(t)}{\sum_{j=1}^{k} n_{ij}(t)}[1+\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*)]^{-1}. \tag{12}$$

Since $\exp(\boldsymbol{x}_i^{'}(t)\beta_i^*), n_{ij}(t) \geq 0$, the non-negativity constraints (4) are met automatically. Combining Eqs. (10) and (12), we obtain Proposition 1.

By Proposition 1, we can estimate the wastage rates when the differential variables of the system are varied. The NHETM is obtained as $\mathbf{Q}(t)\Big|_{\mathbf{x}'(t)} = \left(\hat{q}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} \right)_{i,j \in S}$ with $\hat{q}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} = \hat{p}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} + \hat{p}_{i0}(t)\Big|_{\boldsymbol{x}_i'(t)} \hat{p}_{0j}(t), i, j \in S$, where $\hat{p}_{0j}(t)$ is the estimated entry probability into level j in period t. The entries, $\hat{q}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)}$, are such that $\sum_{j=1}^{k} \hat{q}_{ij}(t)\Big|_{\boldsymbol{x}_i'(t)} = 1$. Following

Osagiede and Ekhosuehi [7] that total enrolment stock expands deterministically at a growth rate, g, the growth rate is estimated in matrix form as

$$\hat{g} = \exp\left(\begin{bmatrix} 0 & 1 \end{bmatrix} \left(\begin{bmatrix} \mathbf{e}^{'} \\ \Phi \end{bmatrix} \begin{bmatrix} \mathbf{e}^{'} & \Phi \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{e}^{'} \\ \Phi \end{bmatrix} \mathbf{N} \right) - 1, \tag{13}$$

where Φ is a $T \times 1$ vector of the time epochs, $\mathbf{N}$ is a $T \times 1$ vector with its entry being the natural logarithm of the total enrolment stock at each time epoch, and $\mathbf{e}^{'}$ is a $T \times 1$ vector of ones. Afterwards, we construct a transition model analogous to the one in [10] for the observed scenario as

$$\bar{\mathbf{q}}(t+1) = \mathbf{q}(0) \prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)}, \tag{14}$$

where $\bar{\mathbf{q}}(t+1)$ denotes the expected relative structure of the system for period $t+1$, $\mathbf{q}(0)$ is the actual relative structure of the system at the base period, $\mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)} = (1+\hat{g})^{-1} \left(\mathbf{Q}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)} + \hat{g}\mathbf{I}\mathbf{1}^{'}\mathbf{P}_0(\varsigma) \right)$ is a $k \times k$ discrete-time qNHIMC, $\mathbf{I}$ is a $k \times k$ identity matrix, $\mathbf{P}_0(\varsigma)$ is a $1 \times k$ entry probability vector in period ς and $\mathbf{1}^{'}$ is a $k \times 1$ vector of ones. The model in Eq. (14) is not used to extrapolate the long-term shift in the structure of the system because $\mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)}$ depends on the NHETM. Thus we propose the following:

Proposition 2: *Suppose the operational mechanism of a $k-$ echelon educational system is such that: (i) the growth rate is deterministic; (ii) the admission is done to replace leavers and to achieve the desired growth; (iii) the existing trend in total enrolment stock is maintained at a point which does not exceed the carrying capacity of each level of the system in the long-run. Then the expected long-term shift in enrolment stocks, $\bar{\boldsymbol{n}}(\tau_v)$, for the period τ_v is*

$$\bar{\boldsymbol{n}}(\tau_v) \sim min\left(\boldsymbol{c}(\tau_v), ceil\left((1+\hat{g})^t \boldsymbol{q}(t) \exp((\tau_v - t)\overline{\boldsymbol{G}}) \times \right.\right.$$

$$\left.\left. \exp\left(\begin{bmatrix} 1 & 0 \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{e}^{'} \\ \Phi \end{bmatrix} \begin{bmatrix} \boldsymbol{e}^{'} & \Phi \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{e}^{'} \\ \Phi \end{bmatrix} \boldsymbol{N} \right) \right) \right),$$

$v = 1, 2, \cdots$, *where $\boldsymbol{c}(\tau_v)$ is the carrying capacity of the system at epoch τ_v.*

Proof: Conditions (i) and (ii) in Proposition 2 have been employed in deriving the matrix $\mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)}$. To prove Proposition 2, we find a stationary continuous-time imbedded Markov chain which is as close as the qNHIMC, $\mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)}$, within some error distance. Equivalently, we solve **P2**. We rewrite the intensity matrix $\mathbf{G}$ in **P2** as $\mathbf{G} = \frac{1}{t+1} \log \prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(t)}$, and then apply the diagonal adjustment method [2]. The solution to P2 is not unique [2]. Suppose there are α possible numbers of $\mathbf{G}$ denoted as $\mathbf{G}^{\alpha}$, $\alpha = 1, 2, \cdots$. Then, the Markov structure

we seek is the one which satisfies:

$$\zeta = min_{\alpha} \left\| \exp(\mathbf{G}^{\alpha}) - \left(\prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)} \right)^{\frac{1}{t+1}} \right\|. \tag{15}$$

Let $\overline{\mathbf{G}}$ be the intensity matrix satisfying Eq. (15). Then the stationary continuous-time imbedded Markov chain is $\sigma^* = \exp(\overline{\mathbf{G}})$. The expression $\sigma^* = \exp(\overline{\mathbf{G}})$ is in line with [1]. By replacing $\prod_{\varsigma=0}^{t} \mathbf{S}(\varsigma) \Big|_{\mathbf{x}'(\varsigma)}$ in Eq. (14) by $\exp((t+1)\overline{\mathbf{G}})$ and in general, by $\exp((\tau_v - t)\overline{\mathbf{G}}), v = 1, 2, \cdots$, we obtain the relation

$$\bar{\mathbf{q}}(\tau_v) \sim \mathbf{q}(t) \exp((\tau_v - t)\overline{\mathbf{G}}), \tau_v > t. \tag{16}$$

By condition (iii) and relation (16) subject to the carrying capacity of the system, the expected enrolment structure over time is obtained as in Proposition 2. The transition model in Proposition 2 is referred to as the neo-SIM model.

§3. Numerical illustration

We demonstrate the utility of the neo-SIM model for equidistant periods using enrolment data for a part-time undergraduate programme as in Table 1 at discrete points and compare the results with the SIM model. All computations are done in Matlab environment. The entries $n_i(t)$ in Table 1 are students enrolment stocks for year t while the other cell entries are students flows. In the six-year graded part-time undergraduate programme, there is neither repetition (except in Year 6) nor demotion. This explains why the diagonal elements are zero for the first five grades in Table 1. We code the sessions $t = 2003/2004, \cdots, 2008/2009$ as $t = 0, \cdots, 5$. Looking at the data in Table 1, we find that the flows during the time period $(t-1, t)$ satisfy: $\sum_{j=1}^{k} n_{ij}(t) = n_{i,i+1}(t), n_{i-1,i}(t-1) \geq n_{i,i+1}(t)$, for $i \in S - \{2, 6\}$, and $\sum_{j=1}^{k} n_{ji}(t) = 0$, for $j > i$. The relation $j > i$ means a demotion, so that $\sum_{j=1}^{k} n_{ji}(t) = 0$, for $j > i$, captures the absence of demotion in the system.

By the method described in [7], we obtain the homogeneous transition matrix from Table 1. Entries in the matrix provide information on the direct transition between levels in the academic programme and the part of wastage replaced by new entrants into the programme. Next, we consider the differential variables of the programme. The tuition fees and charges (i.e., school fees) varied during the period under consideration. For this reason, we use school fees as the explanatory variable in the logistic constraint. We collate records on school fees from the Bursary Department of the University of Benin (Table 2). Using the information in Table 1 and Table 2, we estimate the parameters for the binomial logistic model. The estimates, β_{2i}^* for $i = 1, \cdots, 6$, are not significantly different from zero. However, estimation of the wastage probabilities from Eq. (10) is feasible as the information matrices are non-degenerate. By Propositon 1, we estimate the wastage probabilities for Year $1 - 6$ as shown in Table 3. After that, we calculate the transition probabilities. Using the growth rate estimator in Eq. (13), we obtain the value $\hat{g} = 0.4056$ as the estimated growth rate in

$i \rightarrow j$	1	2	3	4	5	6	$n_{i0}(t)$	$n_i(t)$
$n_{0j}(0)$	112	4	0	0	0	0	-	116
1	0	112	0	0	0	0	0	112
2	0	0	53	0	0	0	0	53
3	0	0	0	56	0	0	0	56
4	0	0	0	0	30	0	0	30
5	0	0	0	0	0	35	0	35
6	0	0	0	0	0	8	10	18
$n_{0j}(1)$	110	0	0	0	0	0	-	110
1	0	106	0	0	0	0	4	110
2	0	0	90	0	0	0	22	112
3	0	0	0	45	0	0	8	53
4	0	0	0	0	48	0	8	56
5	0	0	0	0	0	26	4	30
6	0	0	0	0	0	8	35	43
$n_{0j}(2)$	236	0	0	0	0	0	-	236
1	0	234	0	0	0	0	2	236
2	0	0	78	0	0	0	28	106
3	0	0	0	87	0	0	3	90
4	0	0	0	0	45	0	0	45
5	0	0	0	0	0	43	5	48
6	0	0	0	0	0	13	21	34
$n_{0j}(3)$	353	0	0	0	0	0	-	353
1	0	346	0	0	0	0	7	353
2	0	0	226	0	0	0	8	234
3	0	0	0	78	0	0	0	78
4	0	0	0	0	87	0	0	87
5	0	0	0	0	0	43	2	45
6	0	0	0	0	0	20	36	56
$n_{0j}(4)$	471	180	0	0	0	0	-	651
1	0	470	0	0	0	0	1	471
2	0	0	404	0	0	0	2	406
3	0	0	0	211	0	0	15	226
4	0	0	0	0	78	0	0	78
5	0	0	0	0	0	80	7	87
6	0	0	0	0	0	35	28	63
$n_{0j}(5)$	181	22	0	0	0	0	-	203
1	0	179	0	0	0	0	2	181
2	0	0	489	0	0	0	3	492
3	0	0	0	397	0	0	7	404
4	0	0	0	0	205	0	6	211
5	0	0	0	0	0	67	11	78
6	0	0	0	0	0	44	71	115

Table 1: Enrolment data from 2003/2004 - 2008/2009 at the end of each academic session for B.Sc. Statistics with Computer Science

enrolment stock over the period. The value of $\hat{g}$ is an indication that the expansion in enrolment stock is about 40.56 %. Next, we estimate the qNHIMC, $\mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)}$, for each session $\varsigma = 0, \cdots, 5$. The entries in $\mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)}$ provide information on the probability that losses in the system would result to a consequential admission and that admission would contribute to the desired expansion of the system given the deterministic growth factor. We express $\mathbf{G}$ as $\mathbf{G} = \frac{1}{6}\log\prod_{\varsigma=0}^{5}\mathbf{S}(\varsigma)\Big|_{\mathbf{x}'(t)}$, and then apply the diagonal adjustment regularization method. We obtain the stationary continuous-time imbedded Markov chain σ^* as:

$$\sigma^* = \exp\begin{bmatrix} -0.3755 & 0.1230 & 0.0788 & 0.0443 & 0.0276 & 0.1018 \\ 0.2029 & -0.4434 & 0.1459 & 0.0425 & 0.0291 & 0.0230 \\ 0.1502 & 0.1445 & -0.3863 & 0.0632 & 0.0284 & 0 \\ 0.1368 & 0.0951 & 0.1515 & -0.4291 & 0.0457 & 0 \\ 0.1347 & 0.0910 & 0.1111 & 0.1186 & -0.4554 & 0 \\ 0.1571 & 0.0880 & 0.0930 & 0.0782 & 0.1061 & -0.5224 \end{bmatrix},$$

with $\zeta = 0.0172$. The exact entries in the transition matrix σ^* are computed using the Matlab code *expm()*. By so doing, the transition matrix σ^* is stochastic.

Session	Total fees for returning students	Total fees for new students
2003/2004	28,700	38,900
2004/2005	28,700	38,900
2005/2006	28,700	44,700
2006/2007	28,700	44,700
2007/2008	28,700	60,500
2008/2009	29,700	71,500

Table 2: School fees for B.Sc. Statistics with Computer Science (in Naira)

t	0	1	2	3	4	5
Year 1	0.0185	0.0185	0.0158	0.0158	0.0102	0.0075
Year 2	0.0746	0.0920	0.0920	0.0920	0.0025	0.0415
Year 3	0.0518	0.0518	0.0518	0.0518	0.0518	0.0186
Year 4	0.0276	0.0276	0.0276	0.0276	0.0276	0.0289
Year 5	0.0735	0.0735	0.0735	0.0735	0.0735	0.1410
Year 6	0.6075	0.6075	0.6075	0.6075	0.6075	0.6174

Table 3: Estimates of wastage probabilities for the NHETM

We compare the efficiency of the information transmitted in the use of the neo-SIM model with that of the SIM model for long-term projection using the Shannon entropy rate [11]. We do this because information is closely associated with uncertainty [12]. Earlier, Lee, et al. [13] reported that: when characterizing some unknown events with a statistical model, we should always choose the one that has maximum entropy. In this light, we base our decision

τ_v	τ_1	τ_2	τ_3	τ_5	τ_6	τ_8	τ_{10}
SIM model	0.1871	0.3412	0.4829	0.7331	0.8094	0.8692	0.9161
Neo-SIM model	0.6027	0.7940	0.8704	0.9145	0.9197	0.9226	0.9231

Table 4: Entropy values of the modelled system

on the preferrable model. The results from the entropy rate are presented in Table 4. Table 4 shows that the neo-SIM model is preferrable as it has higher entropy values than the SIM model. Nevertheless, the information is not complete as the entropy values are less than one.

§4. Conclusion

In this study, we have proposed a neo-SIM model for the multi-echelon educational system. The neo-SIM model removes the ill-posed problem arising from raising a homogeneous transition matrix to a fractional index due to distortions and extensions to academic calendar. The major accomplishments of this paper are formalized in Propositions 1 and 2. We illustrate the use of the neo-SIM model with data from a university setting and quantify the information content using the Shannon entropy rate.

Acknowledgement

The authors are very grateful to the anonymous referees for their valuable comments and suggestions to improve the paper in the present form.

References

[1] ISRAEL, R. B., ROSENTHAL, J. S., AND WEI, J. Z. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance Vol. 11*, 2 (2001), 245–265.

[2] KREININ, A. AND SIDELNIKOVA, M. Regularization algorithms for transition matrices. *Algo Research Quarterly Vol. 4*, 1/2 (2001), 23–40.

[3] WOODWARD, M. On forecasting grade, age and length of service distributions in manpower systems. *Journal of the Royal Statistical Society. Series A (General) Vol. 146*, 1 (1983), 74–84.

[4] LINDGREN, B. W. *Statistical theory (4th Ed.)*. Chapman & Hall, New York, 1993.

[5] TSAKLIDIS, G. M. The evolution of the attainable structures of a homogeneous Markov system with fixed size. *Journal of Applied Probability Vol. 31*, 2 (1994), 348–361.

[6] VASSILIOU, P. -C. G. AND GEORGIOU, A. C. Asymptotically attainable structures in nonhomogeneous Markov systems. *Operations Research Vol. 38*, 3 (1990), 537–545.

[7] OSAGIEDE, A. A. AND EKHOSUEHI, V. U. Markovian approach to school enrolment projection process. *Global Journal of Mathematical Sciences Vol. 5*, 1 (2006), 1–7.

[8] MOULD, G. I. Case study of manpower planning for clerical operations. *Journal of the Operational Research Society Vol. 47*, 3 (1996), 358–368.

[9] HARDLE, W., MULLER, M., SPERLICH S., AND WERWATZ, A. *Nonparametric and semiparametric models*. Springer-Verlag, Berlin Heidelberg, 2004.

[10] FEICHTINGER, G. AND MEHLMANN, A. The recruitment trajectory corresponding to particular stock sequences in Markovian person-flow models. *Mathematics of Operations Research Vol. 1*, 2 (1976), 175–184.

[11] GIRARDIN, V. AND LIMNIOS, N. On the entropy for semi-Markov processes. *Journal of Applied Probability Vol. 40*, (2003), 1060–1068.

[12] SHARMA, J. K. *Operations research: theory and applications (4th Ed.)*. Macmillan Publishers, India Ltd., New Delhi, 2009.

[13] LEE, S., VONTA, I., AND KARAGRIGORIOU, A. A maximum entropy type test of fit. *Computational Statistics and Data Analysis Vol. 55*, 9 (2011), 2635–2643.

Virtue U. Ekhosuehi and Augustine A. Osagiede
Department of Mathematics
University of Benin
Benin City, Nigeria.
virtue.ekhosuehi@uniben.edu and augustine.osagiede@uniben.edu

Monografías Matemáticas García de Galdeano **38**, 75–84 (2013)

LINEAR COMBINATION OF BIOMARKERS TO IMPROVE DIAGNOSTIC ACCURACY IN PROSTATE CANCER

Luis Mariano Esteban, Gerardo Sanz and Ángel Borque

Abstract. The combination of multiple biomarkers in order to improve diagnostic accuracy is an important issue in Medicine. Providing an optimal solution to this problem is a widely analyzed issue that does not have a global answer. In different papers, linear combinations of markers that maximize the Area under the Receiver Operating characteristic Curve (ROC) have been proposed. However, none of them can be applied in all possible scenarios. Under the multivariate normal assumption, the best linear combination of markers has been determined, but this hypothesis is not easy to verify in medical data. In this work, we analyze the performance of two non-parametric methods, a step by step algorithm and the min-max combination that have been recently proposed, in order to improve diagnostic accuracy in prostate cancer.

Keywords: biomarkers, ROC curve, linear combinations.

AMS classification: 62P10,62J99.

§1. Introduction

Different variables, such as Prostate Specific Antigen (PSA) in the diagnosis of prostate cancer, have been used to predict a binary clinical outcome. Although some of these markers have a reasonably good ability to discriminate between the categories of an outcome, one single biomarker lacks both the sensitivity and specificity to accurately diagnose prostate cancer or biochemical recurrence after surgery. It is the combination of these biomarkers which may lead to improved overall diagnostic accuracy.

The Receiver operating characteristic (ROC) curve plays a key role in the prediction or prognosis of a binary response. The Area Under the ROC curve (AUC) is the most commonly used parameter to assess a classification model [15], specially, if any of the biomarkers used to build the prediction model is measured on a continuous scale.

Assuming the diagnostic model provides higher probability values for diseased subjects, and choosing a threshold c, we can classify the patients with probability values over c as diseased, and the rest as non-diseased. With this threshold, the sensitivity of a diagnostic test is the proportion of subjects predicted as diseased in the group of actual diseased subjects and the specificity is the proportion of predicted non-diseased subjects in the actual non-diseased group.

The ROC curve is a plot of sensitivity versus 1-Specificity for different threshold probabilities c. In most cases, the area under the ROC curve range from 0.5 to 1, the 0.5 value

corresponds to chance and 1 to perfect accuracy. The simplicity in the interpretation of the AUC is a factor that has helped in its generalization as a measure of the performance of a predictive model.

In recent years, new approaches that improve the ability to predict a disease have been developed. After potential models are built, ROC analysis and the AUC have been a useful tool to select and evaluate the best model [1]. Among others, logistic regression, classification trees or neural networks have been used in classification problems. In those models, cross-validation and bootstrap strategies can be used to estimate AUC, analyzing not only the performance, but also the generalizability of the predictive models [3].

Moreover, the problem of finding combinations of diagnostic tests that maximize the AUC have been extensively addressed in the literature.

Under normality assumption, the best linear combination was provided by Su and Liu [16]. When the multivariate normal distribution of diseased and non-diseased population is assumed,

$$\mathbf{X} \sim N(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}}), \quad \mathbf{Y} \sim N(\mu_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$$

the area under the ROC curve of the optimal linear combination is

$$\mathrm{AUC}_{\max} = \Phi\left(\sqrt{\mu^{\mathrm{T}}(\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}})^{-1}\mu}\right)$$

where Φ denotes the distribution function of the Normal distribution, and the coefficients for the best linear combination are

$$\beta_{max} \propto (\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}})^{-1}\mu$$

where $\mu = \mu_{\mathbf{Y}} - \mu_{\mathbf{X}}$.

The main drawback of these results is the difficulty to verify the Multivariate Normal assumption in many situations. In this context, Pepe and Thompson [11] have proposed non-parametric methods to estimate the linear combination of markers that maximizes the empirical AUC. As the empirical form of the AUC is a step function, an extensive search is required for the optimization purpose and it is computationally intractable when the number of markers is large.

Ma and Huang [9] and Wang et al. [17] have approximated the empirical AUC by a sigmoid function and use this continuous form of the AUC to estimate the best linear combination of variables. As Komori et al. [5] have pointed out, that method followed a rule of thumb to determine a scale parameter and as a result of it the accuracy of the approximation of the empirical AUC is already fixed before running the algorithm. Komori et al. have proposed a method based on a boosting algorithm for maximization of the AUC. They used cross-validation techniques to select the best model. None of those methods is fully non-parametric, and their performance relies on a good selection of some parameters.

However, when the assumed models do not fit the data well, these methods may render invalid and misleading results. Moreover, there are methods that propose the selection of variables based on the optimization of the AUC [7], but the parameters of the models are not estimated via AUC optimization.

Without making assumption on the distribution of the variables (distribution free approach), a step by step algorithm can be a solution computationally less demanding than

an extensive search for the estimation of the parameters of a linear model. Esteban and Sanz [2] analyzed this method through simulation data and a prostate cancer database with satisfactory results. In a similar line, Nicolosi et al. [10] use this non parametric approach to combining markers in the diagnosis of breast cancer.

The estimation of parameters via AUC optimization renders linear models with better discrimination ability than generalized linear models when the true link function is not verified [2, 12], but in most cases the AUC derived from both models is equivalent.

Also, Liu et al. [6] have proposed a min-max combination of biomarkers to improve diagnostic accuracy. This is a linear combination that is not really based on the combination of markers, but rather on the combination of the maximum and minimum of the markers. Since this combination is not a linear model, it can work well in cases where the linearity is not verified by the model.

The purpose of this work is to analyze the performance of two non parametric methods, the step by step algorithm [2] and the min-max combination [6] of biomarkers in order to improve diagnostic accuracy in predictive models of prostate cancer.

§2. The step by step algorithm and the min-max combination

In this section we provide a brief description of the algorithms.

2.1. Step by step algorithm

We consider k biomarkers whose levels are denoted by a vector

$$\mathbf{M} = (M_1, \ldots, M_k)$$

in order to predict a binary outcome D.

The purpose of the step by step algorithm is to estimate the parameters $(\beta_2, \ldots, \beta_k)$ in the linear combination

$$L(M) = M_1 + \beta_2 \cdot M_2 + \ldots + \beta_k \cdot M_k$$

that correspond to maximum AUC. To this end, the method follows $n-1$ steps.

The first step of the approach requires to select the two markers (M_i, M_j), and its combination $M_i + \beta_j M_j$ that corresponds to the maximum AUC. In this extensive search, for each possible combination, the AUC is estimated using the Wilcoxon-Mann-Whitney U statistics.

If we assume that we have n_D observations truly classified as diseased patients, $D = 1$, and $n_{\bar{D}}$ observations corresponding to non-diseased patients, $D = 0$, we can write the results of the test as $\mathbf{M}_{D_1}, \ldots, \mathbf{M}_{D_{n_D}}$ and $\mathbf{M}_{\bar{D}1}, \ldots, \mathbf{M}_{\bar{D}n_{\bar{D}}}$. The Wilcoxon-Mann-Whitney U statistic is given by

$$\widehat{AUC} = \frac{\sum_{i=1}^{n_D} \sum_{j=1}^{n_{\bar{D}}} I(L(\mathbf{M}_{Di}) > L(\mathbf{M}_{\bar{D}j})) + \frac{1}{2} I(L(\mathbf{M}_{Di}) = L(\mathbf{M}_{\bar{D}j}))}{n_D \cdot n_{\bar{D}}}$$

Although β must cover a range in $(-\infty, +\infty)$, selecting β in $[-1, 1]$ provides all possible combinations of (M_i, M_j), because the AUC in $(M_i + \beta \cdot M_j)$ for $\beta < -1$ and $\beta > +1$ is

the same as for $\alpha \cdot M_i + M_j$ where $\alpha = \frac{1}{\beta} \in [-1, +1]$. Choosing 201 equally spaced values of β in $[-1, 1]$ can be a good selection.

Once the combination of two variables that maximizes area in the first step of the algorithm has been calculated , in the second step the linear combination $L_1(M) = M_i + \beta_j M_j$ is considered as a single variable, and a new variable M_p is selected in such a way that their linear combination $L_2(M) = M_i + \beta_j M_j + \beta_p M_p$ corresponds to the maximum AUC.

This process is repeated until all variables are included in the model, thus achieving the best linear combination of markers at each step, in the sense that each model provides the maximum area under the ROC curve.

Note that some aspects of the algorithm are important for its performance, such as the normalization of the input variables or the occurrence of different models with the same area in intermediate steps of the algorithm; for additional details see [2].

2.2. The min-max combination

If we consider again k biomarkers denoted by a vector

$$M = (M_1, \ldots, M_k)$$

we can define

$$M_{\max} = \max_{1 \le i \le k} \{M_i\} \qquad \text{and} \qquad M_{\min} = \min_{1 \le i \le k} \{M_i\}$$

The idea of the procedure is to consider the $M_{\max}$ and the $M_{\min}$ as predictor variables instead of the original $M = (M_1, \ldots, M_k)$. Thus the goal of the algorithm is to estimate the parameter β such that the combination

$$M_\beta = M_{\max} + \beta \cdot M_{\min}$$

maximizes the AUC.

Conditioning on the binary status D, 1 being diseased and 0 otherwise, the biomarkers's levels are denoted by $X = (X_1, \ldots, X_k)$ for a non-diseased subject ($D = 0$) and by $Y = (Y_1, \ldots, Y_k)$ for a diseased subject ($D = 1$). The optimal min-max combination is given by β_{opt} that maximizes the AUC, that is

$$AUC(\beta_{\text{opt}}) = \max_{\beta} Pr\{(Y_{\max} - X_{\max}) + \beta(Y_{\min} - X_{\min}) > 0\}$$

The value of β can be found numerically. Here we calculate it by extensive search using the first step of the step by step algorithm described in the previous section. For more details see [6].

§3. Prostate cancer Data

Prostate cancer (PCa) continues to be the most prevalent solid tumor in men in developed countries. Unfortunately, the discrimination of current predictive tools to predict prostate

cancer or different aspects of PCa is imperfect. In this scenario most investigators agree that prediction models of prostate cancer should be improved by the incorporation of novel biomarkers. [4]

The Prostate Specific Antigen (PSA) is the most commonly used variable in predictive models of prostate cancer [8] and it presents a clear positive skewness. Although the log transformation is commonly used, in most cases the assumption of multivariate normality is not possible for the set of predictive biomarkers. Therefore, non parametric approaches can be a good alternative to build predictive models in prostate cancer.

Our purpose in this work is to explore the performance of the min-max combination and the step by step method in predictive models of prostate cancer. We include as the basis for all models the results of the variable Preoperative PSA measured in the 621 patients of a database of staging prostate cancer. This predictor variable has an AUC of 0.743. Also, another continuous predictor, the Rate of cylinders affected in the biopsy is included in the database, which has an AUC of 0.763. Table 1 shows a short description of the staging prostate cancer database.

PSA				
	Mean	Median	SD	Q_1–Q_3
All cases (n=621)	15.70	8.64	32.718	5.90–15.10
Organ-confined (OC) (n=369)	8.782	7.200	5.635	5.32–10.63
Non Organ-confined (NOC) (n=252)	25.840	13.800	49.236	7.49–25.62
Rate of cylinder affected				
	Mean	Median	SD	Q_1–Q_3
All cases (n=621)	38.230	33.330	26.169	16.67–50.00
Organ-confined (OC) (n=369)	28.210	25.000	18.368	12.50–37.50
Non organ-confined (NOC) (n=252)	52.900	50.000	28.869	30.00–75.00
SD: Standard deviation, Q_1: First Quartile, Q_3: Third Quartile				

Table 1: Staging prostate cancer database (n=621)

To compare the min-max combination and the step by step linear models, we simulated the presence of different continuous variables as new predictors. The purpose is to compare how these methods can build better predictive models. We have simulated different markers that have been added to Prostate Cancer Antigen and Rate of cylinders affected in order to build predictive models of organ confined disease. To estimate the risk of organ-confined disease in newly diagnosed prostate cancer patients is essential to select the best treatment for them: the use of Active Surveillance, focal therapies, surgical procedures or radiotherapy with or without adjuvant hormone therapy.

In a first battery of simulations, we have explored three different scenarios with the staging prostate cancer database. In the first two cases, the simulated new markers follow a multivariate normal distribution with equal and unequal variance-covariance matrix, and in the third one, we simulate data from multivariate log-normal distributions. Denoting X: Non Organ Confined, Y: Organ Confined, the different cases analyzed are the following:

Scenario 1: $\mu_X = (0.5, 0.7, 1), \mu_Y = (0, 0, 0), \Sigma_X = \Sigma_Y = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$

Scenario 2:

$$\begin{aligned} \mu_X &= (0.5, 0.7, 1) \quad , \mu_Y = (0, 0, 0) \\ \Sigma_X &= \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \end{aligned}$$

Scenario 3:

$$\begin{aligned} \mu_{\log X} &= (0.5, 0.7, 1) \quad , \mu_{\log Y} = (0, 0, 0) \\ \Sigma_{\log X} &= \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}, \Sigma_{\log Y} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \end{aligned}$$

Table 2 shows the results from 1000 simulations for every choice of the parameters. The mean, median and standard deviation of the AUC are calculated in step by step linear model (SLM) and min-max combination (min-max). The 95% Confidence intervals (C.I.) are also provided.

Scenario 1	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.8730	0.8734	0.0113	0.8508-0.8945
min-max	0.8184	0.8194	0.0163	0.7860-0.8493
Scenario 2	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.8657	0.8652	0.0082	0.8505-0.8830
min-max	0.8412	0.8424	0.0143	0.8110-0.8662
Scenario 3	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.8373	0.8372	0.0085	0.8213-0.8548
min-max	0.8410	0.8412	0.0141	0.8126-0.8687

Table 2: Simulation results in the staging prostate cancer database ($n_X = 252, n_Y = 369$, 1000 simulations)

In the simulations performed using the staging prostate cancer database, the step by step model provides models with higher discrimination ability than the min-max combination in scenarios 1 and 2, and very similar in scenario 3. These results show that if we add markers that follows a multivariate normal distribution, the step by step method has higher discrimination ability, while the min-max combination method matches this discrimination ability in the case where we move away from this hypothesis.

We want to emphasize that in scenario 1, using only the PSA and the Rate of cylinder affected as predictor variables, the best linear model obtained with the step by step method has an AUC of 0.8146, which is the same as we obtained with the min-max method using the

five predictor variables. Therefore, the step by step method is clearly superior and it seems that min-max method capture the discrimination ability of some variables, but not all.

As a consequence of these results, we have explored another issue which has been recently addressed in the literature. Pepe et al. [13] have analyzed that for models containing standard risk factors and possessing good discrimination, very large independent associations of a new marker with the outcome are required to result in a meaningful larger AUC. In this context, Pinsky and Zhu [14] have analyzed the role of correlation among markers, they conclude that the addition of variables negatively correlated with the previous ones improves greatly the diagnostic accuracy of predictive models.

We think that when negatively correlated variables are added, the min-max method will estimate models with less predictive ability than the step by step method if all variables have a similar discriminatory ability. Therefore, the addition of novel markers to standard ones is not going to be reflected as a clear increase in AUC despite that are negatively correlated, which is the most favorable case.

We select again PSA and Rate of cylinder affected as predictor variables of organ confined disease, and we simulate two new markers that are negatively correlated with the first two ones. We use the min-max method to estimate the best linear combination of $M_{\max}$ and $M_{\min}$ and the the step by step method to combine the 4 variables.

Results show a mean AUC of 0.9064 for min-max method, and a mean AUC of 0.8970, 0.9162 and 0.9380 for the best combination of 2, 3 and 4 variables using the step by step method. The mean AUC obtained with the first method is between the AUC corresponding to models obtained by the step by step method with 2 and 3 variables, and is far from that obtained with four variables. Although the difference between 0.9380 and 0.9064 may seem small, for such high values of the AUC, it is difficult to get a significant increase in its value. Thus, we verify that min-max method fails to capture all discriminatory ability of the set of predictor variables.

§4. Other simulation results

Although in most predictive models of prostate cancer, the PSA variable appears as one of the most important predictors, we want to extend the study to other cases where all markers are simulated.

Now, in scenarios 4 and 5, we simulate data from multivariate normal and multivariate log-normal distributions. Denoting X: Non organ confined disease, Y: Organ confined disease, the mean vector and variance-covariance matrix are:

Scenario 4:

$$\begin{aligned}
\mu_X &= (0.5, 0.5, 0.5, 0.5, 0.5)\ , \mu_Y = (0, 0, 0, 0, 0) \\
\Sigma_X &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \Sigma_Y = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

Scenario 5:

$$\mu_{\log X} = (0.5, 0.6, 0.7, 0.8, 1) \quad , \mu_{\log Y} = (0, 0, 0, 0, 0)$$

$$\Sigma_{\log X} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}, \Sigma_{\log Y} = \begin{pmatrix} 1.5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2.5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Also, in scenario 6 we use transformations of $N(\mu, \sigma)$ to analyze skewed distributions. More specifically we consider:

Scenario 6:

$$X_i = N(1,1)^{-3}, Y_i = N(0,1), i = 1, \ldots, 5.$$

The results from 1000 simulations are displayed in Table 3 for every choice of the parameters. Again, the mean, median, standard deviation and the the 95% C.I. of mean AUC are provided in the step by step linear model (SLM) and the min-max combination.

Scenario 4	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.7883	0.7890	0.0199	0.7487-0.8266
min-max	0.7569	0.7567	0.0210	0.7165-0.7963
Scenario 5	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.7096	0.7087	0.0187	0.6745-0.7490
min-max	0.9106	0.9112	0.0121	0.8855-0.9331
Scenario 6	AUC Mean	AUC Median	AUC SD	AUC 95% C.I.
SLM	0.7115	0.7115	0.0020	0.6754-0.7530
min-max	0.9337	0.9345	0.0126	0.9085-0.9566

Table 3: Simulation results ($n_X = 250, n_Y = 250$, 1000 simulations)

Results show that in scenarios with a clear departure from multivariate normal assumption (Scenarios 5 and 6), the min-max combination shows a superior ability to discriminate than the SLM model, whereas in models using markers that are near to verify the multivariante normal assumption, the step by step method performs better.

In scenario 4, the data are simulated from a normal multivariate and therefore the best linear combination and its corresponding AUC can be calculated theoretically with the expression provided by Su and Liu [16]. Maximum AUC has a value of 0.7854, which is very similar to the value obtained with step by step method (0.7883), with a minimum bias due to the simulation. By contrast, the min-max method has an AUC (0.7569) away from the maximum value, showing an underestimation.

§5. Conclusions

Without clear information about distributional assumptions of biomarkers, non parametric approaches must be taken into account to build predictive models. These nonparametric

methods can achieve a great capacity for discrimination between the different states of a disease depending on the true relation between markers and disease.

The step by step linear model and the min-max combination appear as good alternatives to build predictive models in medicine. The min-max combination performs better when the data depart from normality, but it can be too dependent on the normalization of the variables, which could be a problem for its application to real databases like the staging prostate cancer database.

Another non parametric approach like the step by step method could give a better alternative. It captures the discriminatory ability of all predictor variables and, as a consequence, in the case of a set of predictor variables that have a similar discriminatory ability, it estimates models with greater predictive ability.

Acknowledgements

We gratefully acknowledge the financial support from the MEC project MTM2010-15972. G. Sanz and L.M. Esteban are members of the research group Modelos Estocásticos (D.G.A.).

References

[1] ALEMAYEHU, D., ZOU, KH.. Applications of ROC Analysis in Medical Research. *Academic Radiology 19*, 12 (2012),1457–1464.

[2] ESTEBAN, L.M., SANZ, G., BORQUE, A. A step-by-step algorithm for combining diagnostic tests. *Journal of Applied Statistics 38*, 5 (2011), 899–911.

[3] HARRELL, FE. *Regression Modeling Strategies*. Springer, 2001.

[4] KATTAN MW. Judging New Markers by Their Ability to Improve Predictive Accuracy. *Journal of the National Cancer Institute 7*, 95(9) (2003), 634 –635.

[5] KOMORI, O. AND EGUCHI, S. A boosting method for maximizing the partial area under the ROC curve. *BMC Bioinformatics 11*, (2010), 314.

[6] LIU C, LIU A, HALABI S. A min-max combination of biomarkers to improve diagnostic accuracy. *Stat Med. 30*, 16 (2011), 2005–2014.

[7] LIU, X., JIN, Z. Item reduction in a scale for screening. *Statist. Med. 26* (2007), 4311–4327.

[8] LUGHEZZANI G, BRIGANTI A, KARAKIEWICZ PI, KATTAN MW, MONTORSI F, SHARIAT SF, ET AL. Predictive and Prognostic Models in Radical Prostatectomy Candidates: A Critical Analysis of the Literature. *European Urology 11*, 58(5) (2010), 687–700.

[9] MA, S., HUANG, J. Combining multiple markers for classification using ROC. *Biometrics 63*, 3 (2007),751–7.

[10] NICOLOSI, S., RUSSO, G., D'ANGELO, I., VICARI, G., GILARDI, M.C., BORASI,G. Combining DCE-MRI and 1H-MRS spectroscopy by distribution free approach results in a high performance marker: Initial study in breast patients. *J. Biomedical Science and Engineering 6*, (2013), 357–364.

[11] PEPE, M.S., AND THOMPSON, M.L. Combining diagnostic test results to increase accuracy. *Biostat 1*, 2 (2000), 123–140

[12] PEPE, M.S., CAI, T. AND LONGTON, G. Combining Predictors for Classification Using The Area under the Receiver Operating Characteristic Curve. *Biometrics 62*, (2006), 221–229.

[13] PEPE, M.S., JANES, H., LONGTON, G., LEISENRING, W., NEWCOMB, P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am. J. Epidemiol. 159*, 9 (2004), 882–890.

[14] PINSKY PF, ZHU, CS. Building Multi-Marker Algorithms for Disease Prediction-The Role of Correlation Among Markers. *Biomarker Insights 6*, (2011), 83–93.

[15] STEYERBERG, E.W., PENCINA, M.J., LINGSMA, H.F., KATTAN, M.W., VICKERS, A.J., VAN CALSTER, B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest. 42*, 2 (2012), 216–228.

[16] SU, J.Q. AND LIU, J.S. Linear combinations or multiple diagnostic markers. *J. Amer. Statist. Assoc. 88*, (1993), 1350–1355.

[17] WANG, Z., CHANG, Y., YING, Z. ET AL. A parsimonious threshold-independent protein feature selection method through the area under receiver operating characteristic curve. *Bioinformatics 23*, 20 (2007), 2788–2794.

Luis Mariano Esteban Escaño
Escuela Universitaria Politécnica La Almunia
Universidad de Zaragoza
Mayor s/n, 50100, La Almunia de Goña Godina, Spain.
lmeste@unizar.es

Gerardo Sanz Saiz
Departamento de Métodos estadísticos
Universidad de Zaragoza
Pedro Cerbuna 12, 50009, Zaragoza, Spain.
gerardo@unizar.es

Angel Borque Fernando
Hospital Universitario Miguel Servet
Universidad de Zaragoza
Paseo Isabel La Católica 1-3, 50009, Zaragoza, Spain.
aborque@comz.org

Monografías Matemáticas García de Galdeano **38**, 85–96 (2013)

RESPONSE-ADAPTIVE DESIGNS BASED ON THE EHRENFEST URN

Arkaitz Galbete, José Antonio Moler and Fernando Plo

Abstract.

In this paper we describe a family of response-adaptive designs based on the Ehrenfest urn model, where the previous responses of patients are used in the allocation of the next patient to a treatment. We study some operating characteristics of these designs, such as the power of the usual inferential tests, the variability of the proportion of allocations, the expected failure rate or the target allocation and we compare them with other well-known response-adaptive designs.

Keywords: Response-adaptive designs, Ehrenfest urn.

AMS classification: 60C05, 60F05.

§1. Introduction

A randomized controlled clinical trial is a statistical experiment to compare the efficacy of a new treatment with respect to a control treatment. The control treatment is the best clinical practice known or a placebo. For a discussion on the importance of randomization in clinical trials and a description of different randomized designs see [13].

Urn models have been used to obtain randomized designs. The general procedure can be described as follows. We assume that patients arrive sequentially to the trial. Let δ_n be the indicator variable that takes value 1 if treatment 1, say, the new treatment, is applied, an 0 if treatment 2, say, the control treatment, is applied. Let $N_{n,1}$ be the number of patients allocated to treatment 1 up to the nth patient. Then, we have

$$N_{n,1} = \sum_{k=1}^{n} \delta_k.$$

The number of patients allocated to treatment 2 will be $N_{n,2} = n - N_{n,1}$.

An urn with balls of two different types (or colors) is provided. Let $(W_{n,1}, W_{n,2})$ be the number of balls of each type at stage n of the procedure. When a new patient arrives, a ball is extracted and the patient is allocated to the treatment associated with the ball's type. The probability distribution of δ_n depends on the past history of the procedure only through the composition of the urn at stage $n-1$, and we have

$$P(\delta_n = 1 | W_{n-1,1}) = \frac{W_{n-1,1}}{2w}. \tag{1}$$

Complete randomization can be regarded as an allocation rule that uses an urn with equal number of balls at each stage, $(W_{n,1}, W_{n,2}) = (w, w)$, $w \geq 1$. Then, the distribution of δ_n does not depend on the past history of the procedure and it is clear that

$$P(\delta_n = 1) = \frac{1}{2}. \tag{2}$$

If the urn is modified using the information obtained so far we have an adaptive design.

The Ehrenfest urn design, see [5], is an adaptive design that uses the previous information in the following way. Initially the urn has w balls of each type, that is, $(W_{0,1}, W_{0,2}) = (w, w)$. When a new patient arrives, a ball is drawn from the urn, the patient is allocated to the treatment associated with its type and a ball of the other type is added to the urn. Observe that the total number of balls remains fixed along the process, and $W_{n,1} + W_{n,2} = 2w$, so that $W_{n,1}$ describes completely the state of the urn. The composition of the urn depends only on the number of times that each treatment has been applied.

When the composition of the urn is modified according to the patients' responses we have a response-driven adaptive design. There is a wide catalogue of this kind of designs, see for instance [9].

In this paper we present a family of response-adaptive designs, based on the Ehrenfest urn model, which use the information given by the patients' responses to modify the composition of the urn, and therefore these responses affect the allocation of future patients to treatments. These designs are studied from a theoretical point of view, paying attention to the process $\{W_{n,1}\}$, which describes the evolution of the urn, and to the process $\{N_{n,1}\}$, which describes the evolution of the allocations. Following the program proposed in [8] for any new design, we also study some of the operating characteristics of these designs, such as the power of inferential tests, the variability of the proportion of allocations, the expected failure rate or the target allocation, and we compare them to other well known response-driven adaptive designs.

§2. Ehrenfest response-adaptive designs

The Ehrenfest urn design was introduced in [5] and was slightly generalized, adding partially reflecting barriers in ν and $2w - \nu$, where $0 < \nu < w$, in [6]. A family of designs that can be seen as a generalization of the Ehrenfest urn model were presented and studied in [1] and [2]. Up to our knowledge, these are all the references that use urn models with a fixed number of balls in the design of clinical trials.

Mean and variance can be calculated, at any stage n of the procedure, for the processes $\{W_{n,1}\}$ and $\{N_{n,1}\}$ associated with the Ehrenfest urn design, see [1]. We have

$$E[W_{n,1}] = w, \qquad Var[W_{n,1}] = \frac{w}{2}\left(1 - (1 - \frac{2}{w})^n\right).$$

and also,

$$E[\frac{N_{n,1}}{n}] = \frac{1}{2}, \qquad Var[\frac{N_{n,1}}{n}] = \frac{w}{8n^2}\left(1 - (1 - \frac{2}{w})^n\right).$$

Observe that the process $\{W_{n,1}\}$ is a time homogeneous Markov Chain with state space $E = \{0,\ 1,\ \ldots,\ 2w\}$ and transition probabilities

$$p_{i,j} = \begin{cases} 1 - \frac{i}{2w}, & j = i+1; \\ \frac{i}{2w}, & j = i-1; \end{cases} \qquad i = 0,\ 1,\ \ldots,\ 2w. \tag{3}$$

This property can be exploited, see [2], to obtain strong laws and central limit theorems. In particular, the following central limit theorem holds.

$$\sqrt{n}\left(\frac{N_{n,1}}{n} - \frac{1}{2}\right) \to N(0, \frac{\sigma^2}{(2w)^2}), \quad [D]$$

where σ^2 can be expressed in terms of the eigenvalues and eigenvectors of the transition matrix of the Markov Chain $\{W_{n,1}\}$.

In this paper we propose three different scenarios inspired in well-known response-driven adaptive designs. We assume that patients arrive sequentially and are allocated according to the type of ball extracted. We put this ball back into the urn. The responses to the treatments applied are used to modify the composition of the urn. We consider binary responses and we denote by $Z_{n,i}$, the indicator function which takes value 1 if the response of the nth patient to treatment i is a success, and 0 if it is a failure. Let p_i be the probability of success of treatment i, and $q_i = 1 - p_i$. To avoid trivial cases, we assume that $p_i \in (0,\ 1)$ for $i = 1,\ 2$. We also assume that these responses are independent of the past. In particular, they are independent of the sequence of past and present allocations.

Scenario 1 (S1 design) In scenario $S1$ we reinforce a treatment when it is a success. This reinforcement rule mimics the Randomly Reinforced Urn model, RRU, studied in [4]. In the RRU model, which uses an unbounded urn, the proportion of patients allocated to the best treatment converges to 1 almost surely, and it converges to a beta distribution $\beta(w,\ w)$ when both treatments perform equally, so that this design is not in the scope of the results in [8].

This reinforcement policy has to be adapted to obtain an urn with a fixed number of balls, $2w$. When the urn is in an interior state, that is, when $W_{n,1} \in E \setminus \{0,\ 2w\}$, if the treatment applied is a success, we add a ball of its type and we remove a ball of the other type. If the treatment is a failure, the composition of the urn remains unchanged. S1 design modifies (3) as follows:

$$p_{i,j} = \begin{cases} p_1 \frac{i}{2w}, & j = i+1; \\ q_1 \frac{i}{2w} + q_2(1 - \frac{i}{2w}), & j = i; \\ p_2(1 - \frac{i}{2w}), & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 1,\ \ldots,\ 2w-1, \tag{4}$$

When $W_{n,1} = 0$ or $W_{n,1} = 2w$, additional rules are needed. If the treatment applied is a success, the urn remains unchanged. If it is a failure, we remove a ball of its type and we add a ball of the other type. This amounts to consider states 0 and $2w$ as semi-reflecting barriers. That is

$$[SB]: \begin{array}{ll} p_{0,0} = p_2, & p_{0,1} = q_2 \\ p_{2w,2w} = p_1, & p_{2w,2w-1} = q_1, \end{array} \tag{5}$$

Scenario 2 (S2 design) In scenario $S2$ a treatment is reinforced if it is a success or if the other treatment is a failure. This is the randomized Play-The-Winner rule, PTW, which has

been profusely studied, see [9]. In the PTW design with an unbounded urn, the proportion of patients allocated in treatment 1 converges to $q_2/(q_1+q_2)$; that is, the ratio of allocations to a treatment converges to its relative risk of failure.

This reinforcement policy has to be adapted to obtain an urn with a fixed number of balls, $2w$. When the urn is in an interior state, $W_{n,1} \in E \setminus \{0,\ 2w\}$, we add a ball of type 1 and remove a ball of type 2 if treatment 1 is applied and it is a success or if treatment 2 is applied and it is a failure; treatment 2 is reinforced in a similar way: if treatment 2 is applied and it is a success or if treatment 1 is applied and it is a failure. $S2$ design modifies (3) as follows:

$$p_{i,j} = \begin{cases} p_1 \frac{i}{2w} + q_2(1-\frac{i}{2w}), & j = i+1; \\ p_2(1-\frac{i}{2w}) + q_1 \frac{i}{2w}, & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 1, \ldots, 2w-1, \tag{6}$$

When $W_{n,1} = 0$ or $W_{n,1} = 2w$, we act as in $S1$:

$$[SB]: \begin{array}{ll} p_{0,0} = p_2, & p_{0,1} = q_2 \\ p_{2w,2w} = p_1, & p_{2w,2w-1} = q_1, \end{array} \tag{7}$$

Scenario 3 (S3 design) In scenario $S3$ a treatment is reinforced if the other treatment is applied and it is a failure. This rule is similar to the Drop-The-Loser rule, DTL, introduced in [10]. The DTL rule has the same allocation limit as the PTW rule; that is, the proportion of patients allocated to a treatment converges to its relative risk of failure.

S3 design modifies (3) as follows. The urn remains unchanged if the treatment applied is a success. If it is a failure, we remove a ball of this type and we add a ball of the other type.

$$p_{i,j} = \begin{cases} q_2(1-\frac{i}{2w}), & j = i+1; \\ p_1 \frac{i}{2w} + p_2(1-\frac{i}{2w}), & j = i; \\ q_1 \frac{i}{2w}, & j = i-1; \\ 0, & \text{otherwise}, \end{cases} \qquad i = 0, \ldots, 2w. \tag{8}$$

Note that, in this scenario, the barrier conditions $[SB]$ implicitly hold. Transition matrix (8) was already studied in [11]. We will refer to $S3$ design as the Klein urn design.

Remark 1. Condition [SB] seems quite logical in the spirit of a clinical trial. If $W_{n,1} = 0$ the urn has $2w$ balls of type 2, and treatment 2 is applied until a failure happens. If $W_{n,1} = 2w$ the urn has $2w$ balls of type 1, and treatment 1 is applied until a failure happens. Absorbent barriers would force to apply the same treatment once the barrier is reached. Reflecting barriers make the allocation of patients arriving after reaching the barrier to be deterministic.

Remark 2. Observe that $\{W_{n,1}\}_{n\in\mathbb{N}}$, in the three scenarios, is a finite, irreducible, aperiodic Markov chain. So that, it is positive recurrent and there exists a stationary distribution if $p_i \in (0,\ 1)$ for $i = 1,\ 2$.

A theoretical study of properties of adaptive designs would be of great interest. But, in general, exact values for the mean and the variance of the number of allocations for each n are difficult to obtain when adaptive designs are applied, due to the complicated correlation structure generated between allocations and observed responses. Comparative studies rely heavily on asymptotical properties or on simulation studies. In [8] it is outlined the importance of checking the accuracy of the asymptotic approximations when these theoretical results are used to compare designs.

The family of Ehrenfest response-adaptive urn designs, with a fixed number of balls, which has been presented here, evolve following a recurrence rule which facilitates the computation of exact values for mean and variance of $\{W_{n,1}\}$ for each value of n. In the section 3 we obtain this kind of results for the three designs. Previously, we recall some theoretical results for birth and death chains and we apply them to the allocations process. In the section 4 we make a study of the degree of fulfillment of ethic, randomness and inferential accuracy goals, with respect to other response driven adaptive designs presented in the literature.

§3. Exact values and asymptotic results for Ehrenfest response-adaptive designs

Observe that the evolution of the designs introduced in section 2 is closely related to the Markov Chain $\{W_{n,1}\}$. Therefore, the theory of Markov Chains will be applied in what follows. To facilitate the exposition, we collect in the following proposition some well known results for Markov chains particularized to the process $\{W_{n,1}\}$.

Proposition 1. *For a finite, irreducible and aperiodic Markov Chain $\{W_{n,1}\}$ with state space E:*

a) there exists a stationary distribution $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$, whatever the initial distribution of the chain is.

b) the following strong law holds

$$\frac{1}{n}\sum_{k=1}^{n} W_{k,1} \to \boldsymbol{\pi}^*, \quad a.s.$$

where $\boldsymbol{\pi}^* := \sum_{i\in E} i\pi_i$ is the mean value of the stationary distribution. Besides, for $m > 1$,

$$\frac{1}{n}\sum_{k=1}^{n} W_{k,1}^m \to \boldsymbol{\pi}_m^*, \quad a.s.$$

where $\boldsymbol{\pi}_m^* := \sum_{i\in E} i^m\pi_i$.

c) the following central limit holds

$$\frac{1}{\sqrt{n}}(\sum_{k=1}^{n} W_{k,1} - n\boldsymbol{\pi}^*) \to N(0,\ \sigma^2)$$

where $\sigma^2 = \lim_{n\to\infty} \frac{1}{n} Var[\sum_{k=1}^{n} W_{k,1}]$.

Asymptotic results for the allocation process $\{N_{n,1}\}$ can also be obtained for the designs introduced in section 2.

Proposition 2. *Consider S1, S2 and S3 designs. Then, we have for the process $\{N_{n,1}\}$ the following results*

a) A strong law holds

$$\frac{N_{n,1}}{n} \to \frac{\boldsymbol{\pi}^*}{2w}, \quad a.s.$$

b) A central limit holds

$$\sqrt{n}(\frac{N_{n,1}}{n} - \frac{\boldsymbol{\pi}^*}{2w}) \to N(0,\ \sigma^2/(4w^2))$$

where $\sigma^2 = \lim_{n\to\infty} \frac{1}{n} Var[\sum_{k=1}^{n} W_{k,1}]$.

Proof. Let $\{\mathcal{F}_n\}$ denote the information associated with the design until stage n. For each n, $E[\delta_n|\mathcal{F}_{n-1}] = \dfrac{W_{n-1,1}}{2w}$. Since $\displaystyle\sum_{k=1}^{n} W_{k,1}$ is unbounded, from Corollary 2.3 in [7] $N_{n,1}$ is also unbounded. Then, from the Levi's extension of the Borel-Cantelli theorem,

$$\lim_{n\to\infty} \frac{N_{n,1}}{\displaystyle\sum_{k=1}^{n} W_{k,1}} = 2w, \quad a.s.$$

the proof follows as in Theorem 1 in [2]. □

The stationary distribution of $\{W_{n,1}\}$ can be obtained explicitly for the $S1$, $S2$ and $S3$ designs.

Proposition 3. *The stationary distribution* $\boldsymbol{\pi}$ *of the Markov Chain* $\{W_{n,1}\}$ *satisfies*
a) for the $S1$ *design,*

$$\frac{\pi_i}{\pi_0} = \frac{q_2}{p_2}\left(\frac{p_1}{p_2}\right)^{i-1} \frac{2w}{\binom{2w-1}{i} i},\ i = 1, \ldots, 2w-1 \tag{9}$$

$$\frac{\pi_{2w}}{\pi_0} = \frac{q_2}{q_1}\left(\frac{p_1}{p_2}\right)^{2w-1} \tag{10}$$

b) for the $S2$ *design,*

$$\frac{\pi_i}{\pi_0} = \frac{\displaystyle\prod_{j=0}^{i-1}(q_2 + (p_1 - q_2)\frac{j}{2w})}{\displaystyle\prod_{j=2w-i}^{2w-1}(q_1 + (p_1 - q_2)\frac{j}{2w})},\ i = 1, \ldots, 2w \tag{11}$$

c) for the $S3$ *design,* $\boldsymbol{\pi}$ *is the probability mass function of a Binomial distribution with parameters* $2w$ *and* $q_2/(q_1 + q_2)$).

Proof. As the probability transition matrix of $\{W_{n,1}\}$ under (4), (6), (8) is a tridiagonal transition matrix, the stationary distribution $\boldsymbol{\pi} = \{\pi_i\}_{i=0,\ldots,2w}$ is easy to obtain by solving the balance equations.

$$\frac{\pi_i}{\pi_0} = \prod_{j=0}^{i-1} \frac{p_{j,j+1}}{p_{j+1,j}},\ i = 1, \cdots, 2w,\ \pi_0 = \frac{1}{1 + \sum_{i=1}^{2w} \frac{\pi_i}{\pi_0}} \tag{12}$$

The result is obtained by putting (4), (6), or (8) along with [SB] in (12). □

Remark 3. The asymptotic distribution of the Klein urn design ($S3$ design) was already obtained in [11], making $p = q_1$ and $p' = q_2$.

Remark 4. From Proposition 3, we observe that when $p_1 = p_2 = p$, the stationary distribution is symmetric for the three scenarios. So that, $\boldsymbol{\pi}^* = w$. For scenario S2, when $p = 0.5$ we have the uniform distribution on the set $\{0,\ 1,\ \ldots, 2w\}$.

The following technical result is useful to establish the behavior of mean and variance of the process $\{W_{n,1}\}$ for each value n.

Lemma 4. *Consider the Markov Chain $\{W_{n,1}\}$ for the $S1$, $S2$ or $S3$ designs. Then, both $E[W_{n,1}]$ and $E[W_{n,1}^2]$ satisfy recurrence equations which we will make explicit below.*

Proof. Once the response of the kth patient is obtained, we denote as I_k^+ the indicator variable of adding one ball of type 1 to the urn and I_k^- the indicator variable of removing one ball of type 1 from the urn. We can write the following recurrence equation

$$W_{n+1,1} = W_{n,1} + I_{n+1}^+ - I_{n+1}^-. \tag{13}$$

Note that

$$\begin{aligned} E[I_{n+1}^+|W_{n,1}] &= \sum_{i=1}^{2w-1} p_{i,i+1} I_{\{W_{n,1}=i\}} + q_2 I_{\{W_{n,1}=0\}}, \\ E[I_{n+1}^-|W_{n,1}] &= \sum_{i=1}^{2w-1} p_{i,i-1} I_{\{W_{n,1}=i\}} + q_1 I_{\{W_{n,1}=2w\}}. \end{aligned} \tag{14}$$

As $W_{0,1} = w$, taking expectations in (13) and using (14), we have

$$E_w[W_{n+1,1}] = E_w[W_{n,1}] + \sum_{i=1}^{2w-1} (p_{i,i+1} - p_{i,i-1}) p_{w,i}^n + q_2 p_{w,0}^n - q_1 p_{w,2w}^n. \tag{15}$$

On the other hand, for $i = 1, \ldots, 2w-1$, and depending on the scenario

$$p_{i,i+1} - p_{i,i-1} = \begin{cases} -p_2 + \dfrac{p_1 + p_2}{2w} i, & \text{for the } S1 \text{ design}, \\ -(p_2 - q_2) + \dfrac{p_1 - q_2}{w} i, & \text{for the } S2 \text{ design}, \\ q_2 - \dfrac{q_1 + q_2}{2w} i, & \text{for the } S3 \text{ design}. \end{cases} \tag{16}$$

So that, $p_{i,i+1} - p_{i,i-1} = a_1 + b_1 i$, where a_1 and b_1 are constants determined in (16) depending on the scenario. Then, (15) becomes:

$$E_w[W_{n+1,1}] = (1 + b_1) E_w[W_{n,1}] + a_1 + (q_2 - a_1) p_{w,0}^n - (q_1 + a_1 + 2w b_1) p_{w,2w}^n, \tag{17}$$

which is the recurrence relation for $E_w[W_{n,1}]$ we were looking for.

Now, from (13) we have

$$W_{n+1,1}^2 = W_{n,1}^2 + I_{n+1}^+ + I_{n+1}^- + 2W_{n,1}(I_{n+1}^+ - I_{n+1}^-), \tag{18}$$

Taking expectations in (19) and using (14) we obtain

$$\begin{aligned} E_w[W_{n+1,1}^2] &= E_w[W_{n,1}^2] + \sum_{i=1}^{2w-1}(p_{i,i+1} + p_{i,i-1})p_{w,i}^n + q_2 p_{w,0}^n + q_1(1-4w)p_{w,2w}^n \\ &\quad +2\sum_{i=1}^{2w-1} i(p_{i,i+1} - p_{i,i-1})p_{w,i}^n \end{aligned} \tag{19}$$

Note that, for $i = 1, \ldots, 2w-1$, and depending on the scenario

$$p_{i,i+1} + p_{i,i-1} = \begin{cases} p_2 + \dfrac{p_1 - p_2}{2w} i, & \text{for the } S1 \text{ design,} \\ 1, & \text{for the } S2 \text{ design,} \\ q_2 + \dfrac{q_1 - q_2}{2w} i, & \text{for the } S3 \text{ design.} \end{cases} \tag{20}$$

So that, $p_{i,i+1} + p_{i,i-1} = a_2 + b_2 i$, where a_2 and b_2 are constants determined in (20) for each scenario. Now we can rewrite (19) as follows

$$\begin{aligned} E[W_{n+1,1}^2] &= (1+2b_1)E[W_{n,1}^2] + a_2 + (b_2 + 2a_1)E[W_{n,1}] \\ &\quad +(q_2 - a_2)p_{w,0}^n \\ &\quad +(q_1 - a_2 - 2b_2 w - 4a_1 w - 8b_1 w^2 - 4wq_1)p_{w,2w}^n \end{aligned} \tag{21}$$

which is the recurrence relation for $E[W_{n,1}^2]$.

□

Remark 5. A closed expression for the solution of (17) is easy to obtain for the S3 design:

$$E[W_{1,n}] = \boldsymbol{\pi}^* + r^n(w - \boldsymbol{\pi}^*),$$

where $r = (1 - \dfrac{q_1 + q_2}{w})$. As $E[\delta_i] = E[W_{1,i-1}]/2w$, we have then

$$E[N_{1,n}] = n\frac{\boldsymbol{\pi}^*}{2w} + \frac{1 - r^n}{1 - r}(\frac{1}{2} - \frac{\boldsymbol{\pi}^*}{2w}). \tag{22}$$

The recurrence equations in Lemma 4 provide an alternative way to obtain explicit expressions of $\boldsymbol{\pi}^*$ for the $S1$ and $S2$ designs. For the Klein urn design ($S3$), we obtain the mean and variance of the Binomial distribution with parameters $2w$ and $q_2/(q_1+q_2)$), as expected.

Proposition 5. *Consider the Markov Chain* $\{W_{n,1}\}$. *Then*

a) for the $S1$ *design we have*

$$\boldsymbol{\pi}^* = \frac{2w}{p_1 + p_2}(p_2 + \pi_{2w} - \pi_0)$$

b) for the $S2$ design we have

$$\boldsymbol{\pi}^* = \frac{w}{q_1 - p_2}(q_2 - p_2 + p_2\pi_0 - p_1\pi_{2w})$$

c) for the $S3$ design we have

$$\boldsymbol{\pi}^* = 2w\frac{q_2}{q_1 + q_2},\ \sigma^{*2} = \frac{2wq_1q_2}{(q_1 + q_2)^2}$$

Proof. From Proposition 1 we have that $E_w[W_{n,1}] \to \boldsymbol{\pi}^*$ and $E_w[W_{n,1}^2] \to \boldsymbol{\pi}_2^*$ as $n \to \infty$.

Taking limits in (17) we have

$$\boldsymbol{\pi}^* = \frac{a_1 + (q_2 - a_1)\pi_0 - (q_1 + a_1 + 2wb_1)\pi_{2w}}{-b_1}$$

and a), b) and c) follow taking the coefficients given in (16).

Taking limits in (21) we have

$$\boldsymbol{\pi}_2^* = \frac{a_2 + (b_2 + 2a_1)\boldsymbol{\pi}^* + (q_2 - a_2)\pi_0 + (q_1 - a_2 - 2b_2w - 4a_1w - 8b_1w^2 - 4wq_1)\pi_{2w}}{-2b_1}$$

and the result follows for the $S3$ design taking the coefficients given in (16) and (20).

□

Remark 6. From Proposition 5 we realize that as stated before, when $p_1 = p_2$ the limit allocation is $1/2$ for the three designs but when $p_1 \neq p_2$, the target allocation of $S1$ and $S2$ only depends on the values of the stationary distribution in the barrier states 0 and $2w$ and success probabilities. The limit allocation of $S3$ is the relative risk of failure, as in the PTW and DTL designs.

§4. A comparative study

In this section we present a comparative simulation study among the $S1$, $S2$ and $S3$ designs, described and studied in sections 2 and 3, and their foster rules, the Randomly Reinforced Urn design (RRU), the Play-The-Winner design (PTW) and the Drop-The-Loser rule (DTL). As a benchmark, we use the comparative study among response-driven adaptive designs for two treatments, binary responses and limit allocation in the interval $(0,\ 1)$ which can be find in Chapter 8 of [9]. There, they state that the drop-the-loser design outperforms the play-the-winner design, and also that it is clearly competitive with the double biased coin design. Besides, it attains the minimum variance for the number of allocations among designs with the same limit allocation.

The three designs introduced in section 2 are response adaptive and we will study its inference properties with the classical test for difference of means. From Proposition 2, the proportion of patients allocated in each treatment converges to a value in the interval (0,

1), moreover, if we consider the sigma-algebra $\mathcal{F}_n = \{\delta_{k+1}, Z_{k,1}, Z_{k,2} : \ k = 1, \ldots, n\}$, conditions in [3] hold and, then, Theorem 3.2 in [12] can be applied. So that, the test statistic:

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{N_{1n}} + \frac{p_2 q_2}{N_{2n}}}}$$

where $\hat{p}_i$ is the success proportion among the patients allocated in treatment i converges to a normal distribution. It will be used to test $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$.

Table 1 is built in the spirit of Table 8.3 in [9]. n represents the number of patients that generate a simulated power of 90% for the test introduced before when the complete randomization design is applied. Then, the simulated power and the mean number of failures (with the standard deviation between brackets) are obtained for the $S1$, $S2$ and $S3$ designs and their foster rules. Observe that the expected number of failures for the S3 design is exactly obtained from (22).

			RRU		S1	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	83	75 (12.4)	17	75 (26.1)
0.9	0.5	48	80	12 (3.9)	52	9 (4.9)
0.7	0.4	108	82	43 (6.6)	41	39 (9.4)
			PTW		S2	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	87	75 (8.9)	20	68 (22.9)
0.9	0.5	48	84	11 (3.1)	39	7.4 (2.9)
0.7	0.4	108	88	45 (5.6)	43	36 (5.7)
			DTL		S3	
p_1	p_2	n	Power	Failures	Power	Failures
0.9	0.8	532	89	73 (7.9)	89	72.04
0.9	0.5	48	86	11 (2.6)	87	11.41
0.7	0.4	108	87	44 (5.4)	87	44.30

Table 1: Simulated power and expected number of failures (standard deviation). 5.000 replications.

Note that $S1$ and $S2$ designs are competitive with, or slightly better than their foster rules, RRU and PTW, from the point of view of the number of failures, but clearly inferior from the point ow view of the power of the test statistic. This loss of power could be foreseen from the results in [8], where it is proven that, given a strong law and a central limit theorem for $N_{n,1}/n$, as were obtained in Proposition 2, the higher the variability of $N_{n,1}$, the smaller is the power of the test statistic. In figure 3 the variability of $N_{n,1}/n$, for $n = 200$, $p_1 = 0.8$ and $p_2 = 0.4$ and 1000 replications, is represented and we can see the high variability of $S1$ and $S2$ designs with respect to their foster designs.

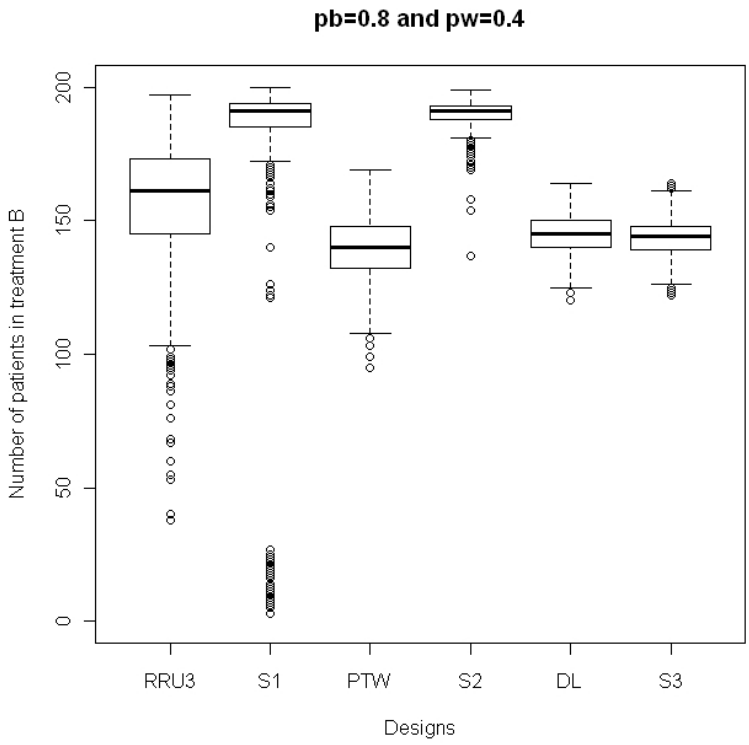


Figure 1: Variability of $N_{n,1}$ for the designs studied.

In this preliminary study, the Klein urn design ($S3$ design) has shown a behavior similar to the DTL rule in number of failures and power (see Table 2) and also in variability of allocations (see Figure 1). Besides, both have the same target allocation. As the Drop-The-Loser rule is perceived as a competitive response adaptive design, see [9], further research on the Klein urn design and its theoretical properties should be encouraged.

Acknowledgements

This research was partially supported by the MEC project MTM2010-15972.

References

[1] BALDI, A. On the speed of convergence of some urn designs for the balanced allocation of two treatments. *Metrika, 62*, 2005, 309–322.

[2] BALDI, A. AND GIANNERINI, S. Generalized Pólya urn designs with null balance. *Journal of Applied Probability, 44*, 2007, 661–669.

[3] BÉLISLE, C. AND MELFI, V. F. Independence after adaptive allocation. *Statistics and Probability Letters, 78*, (2008), 214–224.

[4] MAY, C. AND FLOURNOY, N. Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *The Annals of Statistics, 32*, 2, (2010), 1058–1078.

[5] CHEN, Y.P. Which design is better? Ehrenfest urn versus biased coin. *Advances in Applied Probability, 32*, (2000), 738–749.

[6] CHEN, Y.P. A Central Limit Property under a modified Ehrenfest urn design. *Journal of Applied Probability, 43*, (2006), 409–420.

[7] HALL, P. AND HEYDE, C.C. *Martingale limit theory and its application.*. Academic Press, (1981).

[8] HU, F. AND ROSENBERGER, W. Optimality, variability, power: evaluating response-adaptive ransomization procedures for treatment comparisons. *Journal of the American Statistical Association.* , 98, 463, pp.671-678, (2003).

[9] HU, F. AND ROSENBERGER, W. F. *The theory of response-adaptive randomization in clinical trials*. Wiley, New York, 2006.

[10] IVANOVA, A. A play-the-winner-type urn design with reduced variability. *Metrika*, 58, 2003, 1–13.

[11] KLEIN, M. J. Generalization of the Ehrenfest urn model. *Physical review, 103*, 1, (1956), 17–20.

[12] MELFI, V. F. AND PAGE, C. Estimation after adaptive allocation. *Journal of Statistical Planning and Inference, 87*, (2000), 353–363.

[13] ROSENBERGER, W. F. AND LACHIN, J. M. *Randomization in Clinical Trials. Theory and practice*. Wiley, New York 2002.

Arkaitz Galbete and José Antonio Moler
Department Estadística e Investigación Operativa
University Pública de Navarra
Postal address Campus Arrosadia s/n, 31006, Pamplona, Spain
arkaitz.galbete@unavarra.es, jmoler@unavarra.es

Fernando Plo
Departmento de Métodos Estadísticos and BIFI
Universidad de Zaragoza
Postal address Pedro Cerbuna 12, 50009, Zaragoza, Spain.
fplo@unizar.es

Monografías Matemáticas García de Galdeano **38**, 97–105 (2013)

PHI-DIVERGENCE STATISTICS FOR ORDERED BINOMIAL PROBABILITIES

Nirian Martín, Raquel Mata and Leandro Pardo

Abstract. We consider I independent binomial random variables with parameters n_i and π_i, respectively. In this paper a new family of test statistics based on phi-divergence measures is introduced and studied for the problem of testing hypothesis that involves order constraint on $\{\pi_1, ..., \pi_I\}$. The new family of test statistics contains as a particular case the classical likelihood ratio test.

Keywords: Phi-divergence test statistics, Inequality constrains, Likelihood ratio order, Loglinear models.

AMS classification: 62H17, 62F30.

§1. Introduction

Ordered categorical data with ordered categories appear frequently in the biomedical research literature. For example, in the analysis of a binary response to an increasing exposure data (dose-response experiment). It is well-known that for such data we cannot use the classical chi-square test or likelihood ratio test but we can consider an appropriate analysis that takes into account the ordered categories of a variable as rows of a $I \times 2$ contingency table. Our purpose in this paper is to propose a new family of order-restricted test statistics based on divergence measures that generalize the order-restricted likelihood ratio test as well as the chi-square test.

We consider a modification of an example given in Silvapulle and Sen (2005), in order to motivate the problem considered in this paper. Table 1 contains a subset of data from a prospective study of maternal drinking and congenital malformations. Women completed a questionnaire early in their pregnancy concerning alcohol use in the first trimester; complete data and details are available elsewhere (Graubard and Korn, 1987). Specifically, women were asked what was the amount of alcohol taken during the first three months of their pregnancy and four categories are considered, no alcohol consumption ($i = 1$), average number of alcoholic drinks per day less than one but greater than zero ($i = 2$), one or more and less than three alcoholic drinks per day ($i = 3$) and three or more alcoholic drinks per day ($i = 4$). In terms of a binary outcome, having congenital malformations is considered to be a successful event ($j = 1$).

Let π_i be the probability of a success associated with the i-th alcohol dose. Let us consider some statistical inference questions that may arise in this example and in similar ones with binomial probabilities.

i (drink doses)	n_i	n_{i1} (malformations)	n_{i2} (no malformations)
1 (no drink)	17114	48	17066
2 ($(0,1)$ average drinks)	14502	38	14464
3 ($[1,3)$ average drinks)	793	5	788
$I=4$ (≥ 3 average drinks)	165	2	163

Table 1: Congenital sex-organ malformation relating to maternal alcohol consumption.

1. Is there any evidence of maternal alcohol consumption being related to malformation of sex organ? To answer this question, the null and alternative hypotheses may be formulated as

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 \text{ versus } H_1 : \pi_1, \pi_2, \pi_3, \pi_4 \text{ are not all equal,}$$

respectively. However, this formulation is unlikely to be appropriate because the main issue of interest is the possible increase in the probability of malformation with the increase in alcohol consumption.

2. Is there any evidence that an increase in maternal alcohol consumption is associated with an increase in the probability of malformation?. This question, as it stands, is quite broad to give a precise formulation of the null and the alternative hypotheses. One possibility is to formulate the problem in the following way,

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 \text{ versus } H_1 : \pi_1 \leq \pi_2 \leq \pi_3 \leq \pi_4. \tag{1}$$

Our main purpose in this paper is to present two families of test statistics for testing problems like the problem given in (1). The two families of test statistics are based on phi-divergence measures. The classical likelihood ratio test will be appeared as a particular case. In Section two we present the families of phi-divergence test statistics. Section 3 is devoted to solve the problem presented in this Section. theoretical results.

§2. Phi-divergence test statistics

Consider an experiment with I increasing dose groups. Suppose n individuals are initially placed on experiment, and n_i individuals are assigned to the i-th dose group. The individuals are followed over time for the development of an event of interest. Let N_{i1} be the number of individuals successes, associated with the i-th dose in n_i independent identical trials and N_{i2} the number of no successes, $i = 1, ..., I$. If we denote by π_i the probability of a success associated with the i-th dose, we have that N_{i1} is a Binomial random variable with parameters n_i and π_i, $i = 1, ..., I$, i.e., $N_{i1} \equiv B(n_i, \pi_i)$, $i = 1, ..., I$. The observations obtained can be

displayed in the following way

n_1	n_{11}	$n_{12}=n_1-n_{11}$
$\vdots$	$\vdots$	$\vdots$
n_i	n_{i1}	$n_{i2}=n_i-n_{i1}$
$\vdots$	$\vdots$	$\vdots$
n_I	n_{I1}	$n_{I2}=n_I-n_{I1}$

where here $n_1+n_2+...+n_I=n$ and n_{i1} is the number of successes associated with the Binomial random variable N_{i1}, $i=1,...,I$.

Our interest is in testing

$$H_0:\pi_1=\cdots=\pi_I \text{ versus } H_1:\pi_1\leq\cdots\leq\pi_I. \tag{2}$$

The classical order-restricted likelihood ratio test for testing (2), see for instance Mancuso et al (2001), is given by

$$G^2=2\sum_{i=1}^{I}\left(N_{i1}\log\frac{\widehat{\pi}_i^*}{\widehat{\pi}_0}+(n_i-N_{i1})\log\frac{1-\widehat{\pi}_i^*}{1-\widehat{\pi}_0}\right)$$

being

$$\widehat{\pi}_0=\frac{1}{n}\sum_{i=1}^{I}N_{i1}=\frac{N_{*1}}{n}$$

and $\widehat{\boldsymbol{\pi}}^*=(\widehat{\pi}_1^*,...,\widehat{\pi}_I^*)^T$ the MLE of $\boldsymbol{\pi}=(\pi_1,...,\pi_I)^T$ under the alternative hypothesis. This estimator can be obtained using the PAVA algorithm.

Let $\widehat{\boldsymbol{p}}_j=\left(\frac{N_{1j}}{n_1},...,\frac{N_{Ij}}{n_I}\right)^T$, $j=1,2$. The MLE of $\boldsymbol{\pi}=(\pi_1,...,\pi_I)^T$ is $\widehat{\boldsymbol{p}}_1$. If $\frac{N_{i1}}{n_i}\leq\frac{N_{i+1,1}}{n_{i+1}}$ $\forall i=1,...,I-1$ we have $\widehat{\boldsymbol{\pi}}^*=(\widehat{\pi}_1^*,...,\widehat{\pi}_I^*)^T=\widehat{\boldsymbol{p}}_1$. Otherwise, if there is an index $h\in\{1,...,I-1\}$ such that $\frac{N_{h1}}{n_h}>\frac{N_{h+1,1}}{n_{h+1}}$ the elements $\widehat{p}_h$ and $\widehat{p}_{h+1}$ are called "violators". In this case we replace $\widehat{p}_h$ and $\widehat{p}_{h+1}$ by their weighted average

$$AV_{h,h+1}=\frac{N_{h1}+N_{h+1,1}}{n_h+n_{h+1}}.$$

Then $\widehat{\pi}_h^*=\widehat{\pi}_{h+1}^*=AV_{h,h+1}$. If the new set of $I-1$ verifies

$$\widehat{\pi}_h^*\leq\widehat{\pi}_{h+1}^*,\ h=1,...,I-1,$$

the PAVA optimization problem is finished. Otherwise we iterate the previous process.

We define the $2I$-dimensional probability vectors,

$$\widehat{\boldsymbol{p}}=\left(\left(\left(\bigoplus_{i=1}^{I}\frac{n_i}{n}\right)\widehat{\boldsymbol{p}}_1\right)^T,\left(\left(\bigoplus_{i=1}^{I}\frac{n_i}{n}\right)\widehat{\boldsymbol{p}}_2\right)^T\right)^T=\left(\tfrac{N_{11}}{n},....,\tfrac{N_{I1}}{n},\tfrac{N_{12}}{n},....,\tfrac{N_{I2}}{n}\right)^T$$

$$\boldsymbol{p}(\widehat{\boldsymbol{\theta}}) = \left(\tfrac{n_1}{n}\widehat{\pi}_0, ..., \tfrac{n_I}{n}\widehat{\pi}_0, \tfrac{n_1}{n}(1-\widehat{\pi}_0), ..., \tfrac{n_I}{n}(1-\widehat{\pi}_0)\right)^T =$$

$$\left(\tfrac{n_1}{n}\tfrac{N_{*1}}{n}, ..., \tfrac{n_I}{n}\tfrac{N_{*1}}{n}, \tfrac{n_1}{n}(1-\tfrac{N_{*1}}{n}), ..., \tfrac{n_I}{n}(1-\tfrac{N_{*1}}{n})\right)^T$$

and

$$\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}) = \left(\tfrac{n_1}{n}\widehat{\pi}_1^*, ..., \tfrac{n_I}{n}\widehat{\pi}_I^*, \tfrac{n_1}{n}\left(1-\widehat{\pi}_1^*\right), ..., \tfrac{n_I}{n}\left(1-\widehat{\pi}_I^*\right)\right)^T.$$

It is an easy exercise to verify that

$$G^2 = 2n(\mathrm{d}_{Kull}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})) - \mathrm{d}_{Kull}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widetilde{\boldsymbol{\theta}}))), \tag{3}$$

where $\mathrm{d}_{Kull}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $\mathrm{d}_{Kull}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widetilde{\boldsymbol{\theta}})))$ are the Kullback-Leibler divergence between the $2I$-dimensional probability vectors $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\widehat{\boldsymbol{\theta}})$ in the first case and between $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\widetilde{\boldsymbol{\theta}})$ in the second case. The Kullback-Leibler divergence measure between two $2I$-dimensional probability vectors $\boldsymbol{p} = (p_{11}, ..., p_{I1}, p_{12}, ..., p_{I2})^T$ and $\boldsymbol{q} = (q_{11}, ..., q_{I1}, q_{11}, ..., q_{I2})^T$, is given by

$$\mathrm{d}_{Kull}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{I}\left(p_{i1}\log(\tfrac{p_{i1}}{q_{i1}}) + p_{i2}\log(\tfrac{p_{i2}}{q_{i2}})\right).$$

The classical order-restricted chi-square test statistic for testing (2), known as Bartholomew's test-statistic, is given by

$$X^2 = \frac{1}{\frac{N_{*1}}{n}\left(1-\frac{N_{*1}}{n}\right)}\sum_{i=1}^{I} n_i\left(\widehat{\pi}_i^* - \frac{N_{*1}}{n}\right)^2, \tag{4}$$

and the test statistics X^2 can be written as

$$X^2 = 2n\mathrm{d}_{Pearson}(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}), \boldsymbol{p}(\widehat{\boldsymbol{\theta}})), \tag{5}$$

where $\mathrm{d}_{Pearson}(\boldsymbol{p}, \boldsymbol{q})$ is the Pearson divergence measure defined by

$$\mathrm{d}_{Pearson}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{I}\left(\tfrac{(p_{i1}-q_{i1})^2}{q_{i1}} + \tfrac{(p_{i2}-q_{i2})^2}{q_{i2}}\right).$$

Details about this test-statistic can be found in Fleiss et al. (2003, Section 9.3).

More general than the Kullback-Leibler divergence and Pearson divergence measures are ϕ-divergence measures, defined as

$$d_\phi(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{I}\left(q_{i1}\phi\left(\tfrac{p_{i1}}{q_{i1}}\right) + q_{i2}\phi\left(\tfrac{p_{i2}}{q_{i2}}\right)\right),$$

where $\phi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ is a convex function such that $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$, $0\phi(\frac{0}{0}) = 0$, $0\phi(\frac{p}{0}) = p\lim_{u\to\infty}\frac{\phi(u)}{u}$, for $p \neq 0$. For more details about ϕ-divergence measures see Pardo (2006).

Based on ϕ-divergence measures we shall consider in this paper two families of order-restricted ϕ-divergence test statistics valid for testing (2). The first one generalizes the order-restricted likelihood ratio test given in (3) and its expression is

$$T_\phi(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \frac{2n}{\phi''(1)}(\mathrm{d}_\phi(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) - \mathrm{d}_\phi(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}))) =$$

$$\frac{2}{\phi''(1)}\left\{\sum_{i=1}^{I} n_i \left(\frac{N_{*1}}{n}\phi\left(\frac{\frac{N_{i1}}{n}}{\frac{n_i}{n}\frac{N_{*1}}{n}}\right) + \frac{N_{*2}}{n}\phi\left(\frac{\frac{N_{i2}}{n}}{\frac{n_i}{n}\frac{N_{*2}}{n}}\right) - \widehat{\pi}_i^*\phi\left(\frac{\frac{N_{i1}}{n}}{\frac{n_i}{n}\widehat{\pi}_i^*}\right) - (1-\widehat{\pi}_i^*)\,\phi\left(\frac{\frac{N_{i2}}{n}}{\frac{n_i}{n}\left(1-\widehat{\pi}_i^*\right)}\right)\right)\right\}.$$

For $\phi(x) = x\log x - x + 1$, we get the likelihood ratio test. The second one generalize the order-restricted Pearson test statistic

$$S_\phi(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \frac{2n}{\phi''(1)} d_\phi(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) =$$

$$\frac{2}{\phi''(1)}\left\{\sum_{i=1}^{I} n_i \left(\frac{N_{*1}}{n}\phi\left(\frac{\widehat{\pi}_i^*}{\frac{N_{*1}}{n}}\right) + \frac{N_{*2}}{n}\phi\left(\frac{(1-\widehat{\pi}_i^*)}{\frac{N_{*2}}{n}}\right)\right)\right\}.$$

For $\phi(x) = \frac{1}{2}(x-1)^2$, we get the chi-square distribution.

Under H_0, the asymptotic distribution of $S_\phi(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $T_\phi(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ is

$$\begin{aligned}\lim_{n\to\infty}\Pr\left(S_\phi(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) \le x\right) &= \lim_{n\to\infty}\Pr\left(T_\phi(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}),\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) \le x\right)\\ &= \sum_{i=0}^{I-1} w_i(\boldsymbol{\theta}_0)\Pr\left(\chi^2_{I-1-i} \le x\right)\end{aligned}$$

where $\boldsymbol{\theta}_0$ is the true value and unknown parameter, $\chi^2_0 \equiv 0$. We can observe that the asymptotic distribution does not depend on the function ϕ under consideration. The result has been obtained by following similar steps as in Theorem 1 in Martin, Mata and Pardo (2012) taking $J = 2$.

Since $\boldsymbol{\theta}_0$ is unknown, we cannot use directly the previous result. However, the unknown parameter $\boldsymbol{\theta}_0$ can be replaced by its estimator under the null hypothesis, $\widehat{\boldsymbol{\theta}}$. The tests performed replacing $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}$ are called "local tests" (see Dardanoni and Forcina (1998)) and they are usually considered to be good approximations of the theoretical tests.

It has been established (see, for instance Silvapulle and Sen (2005)) that for $I = 2,3,4$ we have an explicit expression for $w_i(\boldsymbol{\theta}_0)$ on the basis of a variance-covariance matrix

$$\boldsymbol{V}(\boldsymbol{\theta}_0) = \boldsymbol{G}_{(I-1)\times(I-1)}\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{G}^T_{(I-1)\times(I-1)} = (v_{ih}(\boldsymbol{\theta}_0))_{i,h=1,\ldots,I-1}$$

with $\boldsymbol{G}_{(I-1)\times(I-1)}$ being a $(I-1)\times(I-1)$ matrix with 1's in the main diagonal and -1's in their upper diagonal and

$$\mathbf{B}(\boldsymbol{\theta}_0) = \mathcal{I}_{11}^{-1}(\boldsymbol{\theta}_0) + \mathcal{I}_{11}^{-1}(\boldsymbol{\theta}_0)\mathcal{I}_{12}(\boldsymbol{\theta}_0)\left(\mathcal{I}_{22}(\boldsymbol{\theta}_0) - \mathcal{I}_{12}(\boldsymbol{\theta}_0)^T\mathcal{I}_{11}^{-1}(\boldsymbol{\theta}_0)\mathcal{I}_{12}(\boldsymbol{\theta}_0)\right)^{-1}\mathcal{I}_{12}(\boldsymbol{\theta}_0)^T\mathcal{I}_{11}^{-1}(\boldsymbol{\theta}_0)$$

The expressions of $\mathcal{I}_{11}(\boldsymbol{\theta}_0), \mathcal{I}_{12}(\boldsymbol{\theta}_0), \mathcal{I}_{22}(\boldsymbol{\theta}_0)$ and $\mathcal{I}_{12}(\boldsymbol{\theta}_0)$ are obtained through

$$\mathcal{I}_F(\boldsymbol{\theta}) = \boldsymbol{W}^T (D_{\boldsymbol{p}(\boldsymbol{\theta})} - \boldsymbol{p}(\boldsymbol{\theta})\boldsymbol{p}^T(\boldsymbol{\theta}))\, \boldsymbol{W} = \begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}) & \mathcal{I}_{12}(\boldsymbol{\theta}) \\ \mathcal{I}_{21}(\boldsymbol{\theta}) & \mathcal{I}_{22}(\boldsymbol{\theta}) \end{pmatrix}, \tag{6}$$

where $D_{\boldsymbol{a}}$ is the diagonal matrix of vector $\boldsymbol{a}$ and $\boldsymbol{W} = (\boldsymbol{W}_{12}, \boldsymbol{W}_1, \boldsymbol{w}_2)$ is the full rank design matrix of size $2I \times (2I-1)$, such that $\boldsymbol{w}_2^T = (\mathbf{1}_I^T, \mathbf{0}_I^T)$,

$$\boldsymbol{W}_{12}^T = (\boldsymbol{I}_{(I-1)\times(I-1)}, \mathbf{0}_{(I+1)\times(I-1)}),\ \boldsymbol{W}_1^T = (\boldsymbol{I}_{(I-1)\times(I-1)}, \mathbf{0}_{(I-1)}, \boldsymbol{I}_{(I-1)\times(I-1)}, \mathbf{0}_{(I-1)}),$$

$\boldsymbol{I}_{a\times a}$ is the the identity matrix of order a, $\mathbf{0}_a$ is the a-vector of zeros and $u = -\log(\mathbf{1}_{2I}^T \exp\{\boldsymbol{W}\boldsymbol{\theta}\})$.

The explicit expressions of $w_i(\boldsymbol{\theta}_0)$ are as follows:

a) $I = 2$,

$$w_0(\boldsymbol{\theta}_0) = w_1(\boldsymbol{\theta}_0) = 0.5;$$

b) $I = 3$

$$w_0(\boldsymbol{\theta}_0) = 0.5 - w_2(\boldsymbol{\theta}_0),\ w_1(\boldsymbol{\theta}_0) = 0.5,\ w_2(\boldsymbol{\theta}_0) = \frac{1}{2\pi} \arccos\left(\rho_{12}(\boldsymbol{\theta}_0)\right);$$

c) $I = 4$

$$\begin{aligned} w_0(\boldsymbol{\theta}_0) &= \frac{1}{4\pi}\left(2\pi - \arccos\left(\rho_{12}(\boldsymbol{\theta}_0)\right) - \arccos\left(\rho_{13}(\boldsymbol{\theta}_0)\right) - \arccos\left(\rho_{23}(\boldsymbol{\theta}_0)\right)\right), \\ w_1(\boldsymbol{\theta}_0) &= \frac{1}{4\pi}\left(3\pi - \arccos\rho_{12\bullet 3}(\boldsymbol{\theta}_0) - \arccos\rho_{13\bullet 2}(\boldsymbol{\theta}_0) - \arccos\rho_{23\bullet 1}(\boldsymbol{\theta}_0)\right), \\ w_2(\boldsymbol{\theta}_0) &= 0.5 - w_0(\boldsymbol{\theta}_0), \\ w_3(\boldsymbol{\theta}_0) &= 0.5 - w_1(\boldsymbol{\theta}_0), \end{aligned}$$

where

$$\rho_{ih}(\boldsymbol{\theta}_0) = \frac{v_{ih}(\boldsymbol{\theta}_0)}{\sqrt{v_{ii}(\boldsymbol{\theta}_0) v_{hh}(\boldsymbol{\theta}_0)}}$$

is the correlation coefficient between i and h and

$$\rho_{ih\bullet\nu}(\boldsymbol{\theta}_0) = \frac{\rho_{ih}(\boldsymbol{\theta}_0) - \rho_{i\nu}(\boldsymbol{\theta}_0)\rho_{\nu h}(\boldsymbol{\theta}_0)}{\sqrt{\left(1-\rho_{i\nu}^2(\boldsymbol{\theta}_0)\right)\left(1-\rho_{\nu h}^2(\boldsymbol{\theta}_0)\right)}}, \quad \nu \in \{1,2,3\} - \{i,h\}$$

the partial correlation coefficient between i and h given a set of variables with indices in $\{1,2,3\}$.

For a general value of I we can use the Monte-Carlo method. It is worthwhile to mention that these values can be also computed using `mvtnorm` R package (see http://CRAN.R-project.org/package=mvtnorm, for details), however this method based on numerical integration tends to provide less accurate values.

§3. Example

In this section we are going to analyze the first data set of the introduction (Table 1), where $I = 4$. The sample, a realization of $\boldsymbol{N}$, is summarized in vector

$$\begin{aligned}\boldsymbol{n} &= (n_{11}, n_{21}, n_{31}, n_{41}, n_{12}, n_{22}, n_{32}, n_{42})^T \\ &= (48, 38, 5, 2, 17066, 14464, 788, 163)^T.\end{aligned}$$

The estimated probability (sub)vectors of interest are

$$\begin{aligned}\widehat{\boldsymbol{p}}_1 &= \left(\tfrac{48}{17114}, \tfrac{38}{14502}, \tfrac{5}{793}, \tfrac{2}{165}\right)^T = (0.0028, 0.0026, 0.0063, 0.0121)^T, \\ \widehat{\boldsymbol{\pi}}^* &= \left(\tfrac{43}{15\,808}, \tfrac{43}{15\,808}, \tfrac{5}{793}, \tfrac{2}{165}\right)^T = (0.0027, 0.0027, 0.0063, 0.0121)^T, \\ \widehat{\pi}_0 &= \tfrac{93}{32\,574},\end{aligned}$$

and

$$\begin{aligned}\widehat{\boldsymbol{p}} =&(\tfrac{48}{32\,574}, \tfrac{38}{32\,574}, \tfrac{5}{32\,574}, \tfrac{2}{32\,574}, \tfrac{17066}{32\,574}, \tfrac{14464}{32\,574}, \tfrac{788}{32\,574}, \tfrac{163}{32\,574})^T \\ =&(1.4736 \times 10^{-3}, 1.166\,6 \times 10^{-3}, 0.1535 \times 10^{-3}, 0.0614 \times 10^{-3}, \\ &0.5239, 0.4440, 24.191 \times 10^{-3}, 5.004 \times 10^{-3})^T, \\ \boldsymbol{p}(\widetilde{\boldsymbol{\theta}}) =&\left(\tfrac{17114}{32\,574}\tfrac{43}{15\,808}, \tfrac{14502}{32\,574}\tfrac{43}{15\,808}, \tfrac{793}{32\,574}\tfrac{5}{793}, \tfrac{165}{32\,574}\tfrac{2}{165}, \right. \\ &\tfrac{17114}{32\,574}(1 - \tfrac{43}{15\,808}), \tfrac{14502}{32\,574}(1 - \tfrac{43}{15\,808}), \tfrac{793}{32\,574}(1 - \tfrac{5}{793}), \tfrac{165}{32\,574}\left.(1 - \tfrac{2}{165})\right)^T \\ =&(1.4291 \times 10^{-3}, 1.2110 \times 10^{-3}, 0.1535 \times 10^{-3}, 0.0614 \times 10^{-3}, \\ &0.5240, 0.4440, 24.191 \times 10^{-3}, 5.\,004 \times 10^{-3})^T, \\ \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) =&\left(\tfrac{17114}{32\,574}\tfrac{93}{32\,574}, \tfrac{14502}{32\,574}\tfrac{93}{32\,574}, \tfrac{793}{32\,574}\tfrac{93}{32\,574}, \tfrac{165}{32\,574}\tfrac{93}{32\,574}, \right. \\ &\tfrac{17114}{32\,574}(1 - \tfrac{93}{32\,574}), \tfrac{14502}{32\,574}(1 - \tfrac{93}{32\,574}), \tfrac{793}{32\,574}(1 - \tfrac{93}{32\,574}), \tfrac{165}{32\,574}\left.(1 - \tfrac{93}{32\,574})\right)^T, \\ =&(1.5 \times 10^{-3}, 1.2711 \times 10^{-3}, 0.0695 \times 10^{-3}, 0.01446 \times 10^{-3}, \\ &0.5239, 0.443\,9, 24.275 \times 10^{-3}, 5.050\,9 \times 10^{-3})^T.\end{aligned}$$

and the estimators of the weights are obtained through

$$\boldsymbol{V}(\widehat{\boldsymbol{\theta}}) = \begin{pmatrix} 83774.0156250 & -14428.7177734 & 0 \\ -14428.7177734 & 15217.7109375 & -788.9927979 \\ 0 & -788.9927979 & 1457.5666504 \end{pmatrix},$$

that is

$$\begin{aligned}\rho_{12}(\widehat{\boldsymbol{\theta}}) &= \frac{-14428.7177734}{\sqrt{83774.0156250 \times 15217.7109375}} = -0.40411, \\ \rho_{13}(\widehat{\boldsymbol{\theta}}) &= 0, \\ \rho_{23}(\widehat{\boldsymbol{\theta}}) &= \frac{-788.9927979}{\sqrt{15217.7109375 \times 1457.5666504}} = -0.16753.\end{aligned}$$

$$\rho_{12\bullet 3}(\boldsymbol{\theta}_0) = -0.4099,$$
$$\rho_{13\bullet 2}(\boldsymbol{\theta}_0) = -7.\,507\,2 \times 10^{-2},$$
$$\rho_{23\bullet 1}(\boldsymbol{\theta}_0) = -0.183\,15.$$

$$w_0(\widehat{\boldsymbol{\theta}}) = \frac{1}{4\pi}\left(2\pi - \arccos\left(-0.40411\right) - \arccos\left(0\right) - \arccos\left(-0.16753\right)\right) = 0.07850,$$
$$w_1(\widehat{\boldsymbol{\theta}}) = \frac{1}{4\pi}\left(3\pi - \arccos(-0.4099) - \arccos(-7.\,507\,2 \times 10^{-2}) - \arccos(-0.183\,15)\right)$$
$$= 0.32075,$$
$$w_2(\widehat{\boldsymbol{\theta}}) = 0.5 - w_0(\widehat{\boldsymbol{\theta}}) = 0.5 - 0.07850 = 0.4215,$$
$$w_3(\widehat{\boldsymbol{\theta}}) = 0.5 - w_1(\widehat{\boldsymbol{\theta}}) = 0.5 - 0.32075 = 0.17925,$$

$$\{w_i(\widehat{\boldsymbol{\theta}})\}_{i=0}^{I-1} = \{0.0785, 0.32075, 0.4215, 0.17925\}. \tag{7}$$

From these weights the quantile of order 0.05, 5.02, is easy to calculate by following a bisection method.

If we take $\phi_\lambda(x) = \frac{1}{\lambda(1+\lambda)}(x^{\lambda+1} - x - \lambda(x-1))$, where for each $\lambda \in \mathbb{R} - \{-1, 0\}$, "power divergence family of measures" is obtained

$$d_\lambda(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{\lambda(\lambda+1)}\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{p_{ij}^{\lambda+1}}{q_{ij}^{\lambda}} - 1\right), \text{ for each } \lambda \in \mathbb{R} - \{-1, 0\}. \tag{8}$$

It is also possible to cover the real line for λ, by defining $d_\lambda(\boldsymbol{p}, \boldsymbol{q}) = \lim_{t\to\lambda} d_t(\boldsymbol{p}, \boldsymbol{q})$, for $\lambda \in \{-1, 0\}$. It is well known that $d_0(\boldsymbol{p}, \boldsymbol{q}) = d_{Kull}(\boldsymbol{p}, \boldsymbol{q})$ and $d_1(\boldsymbol{p}, \boldsymbol{q}) = d_{Pearson}(\boldsymbol{p}, \boldsymbol{q})$, which is very interesting because the power divergence based family of test-statistics, which contains as special cases G^2 and X^2, can be created. In Table 2, the power divergence based test-statistics and their corresponding asymptotic p-values are shown.

test-statistic	$\lambda=-1.5$	$\lambda=-1$	$\lambda=-0.5$	$\lambda=0$	$\lambda=\frac{2}{3}$	$\lambda=1$	$\lambda=1.5$	$\lambda=2$
$T_\lambda(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widetilde{\boldsymbol{\theta}}), \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$	3.3068	3.8173	4.4920	5.4057	7.2076	8.4895	11.1549	15.1820
$p-\text{value}(T_\lambda)$	0.1177	0.0911	0.0650	0.0413	0.0169	0.0090	0.0024	0.0003
$S_\lambda(\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}), \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$	3.2993	3.8124	4.4896	5.4057	7.2107	8.4942	11.1617	15.1911
$p-\text{value}(S_\lambda)$	0.1181	0.0913	0.0651	0.0413	0.0169	0.0090	0.0024	0.0003

Table 2: Power divergence based test-statistics and asymptotic p-values.

It can be seen that with a significance level equal to 0.05 we accept the null hypothesis that there is no trend in probabilities when $\lambda \in \{-1.5, -1, -0.5\}$, while we cannot accept when $\lambda \in \{0, \frac{2}{3}, 1, 1.5, 2\}$. In particular, we cannot accept lack of trend in binomial probabilities when the test-statistic is the likelihood ratio test-statistic (G^2) or Bartholomew's test-statistic (X^2). A simulation study could be helpful in order to analyze the exact behaviour of the phi-divergence based test-statistics and decide which test-statistic has the best performance.

Acknowledgements

This work is partially supported by the Spanish Ministry of Education and Science, Research Project MTM2009-10072.

References

[1] R.E. BARLOW, D.J. BARTHOLOMEW AND H.D. BRUNK. *Statistical inference under order restrictions*. Wiley, 1972.

[2] V. DARDANONI AND A. FORCINA. A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context. *Journal of the American Statistical Association*, 93 (1998), 1112–1122.

[3] J.L. FLEISS, B. LEVIN AND M.C. PAIK. *Statistical Methods for Rates and Proportions*. Wiley Interscience, 2003.

[4] B. I. GRAUBARD AND E. L. KORN. Choice of Column Scores for Testing Independence in Ordered $2 \times I$ Contingency Tables. *Biometrics*, 43 (1987), 471–476.

[5] N. MARTIN AND L. PARDO. New families of estimators and test statistics in log-linear models. *Journal of Multivariate Analysis*, 99 (8) (2008), 1590–1609.

[6] N. MARTIN, R. MATA AND L. PARDO. Phi-divergence statistics for loglinear models subject to constraints of likelihood ratio order. *Working paper*, Department of Statistics and O.R. (Complutese University of Madrid), (2012).

[7] J.Y. MANCUSO, H. AHAN AND J.J. CHEN. Order-restricted dose-related trend tests. *Statistics in Medicine* , 20 (2001), 2305–2318.

[8] L. PARDO. *Statistical Inference Based on Divergence Measures*. Statistics: series of Textbooks and Monograhps. Chapman & Hall / CRC, 2006.

[9] M.J. SILVAPULLE AND P.K. SEN. *Constrained statistical inference. Inequality, order, and shape restrictions*. Wiley Series in Probability and Statistics. Wiley-Interscience (John Wiley & Sons), 2005.

Nirian Martín
Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid 126, 28903 Getafe (Madrid), Spain
nirian.martin@uc3m.es

Raquel Mata
Departamento de Estadística e I.O. I
Universidad Complutense de Madrid
Plaza de Ciencias 3, 28040 Madrid, Spain
raquel.mata@pdi.ucm.es

Leandro Pardo
Departamento de Estadística e I.O. I
Universidad Complutense de Madrid
Plaza de Ciencias 3, 28040 Madrid, Spain
lpardo@mat.ucm.es

Monografías Matemáticas García de Galdeano **38**, 107–116 (2013)

A DECISION MODEL FOR A NEWSVENDOR INVENTORY PROBLEM WITH AN EXTRAORDINARY ORDER

Valentín Pando, Luis A. San-José, Juan García-Laguna and Joaquín Sicilia

Abstract. This work presents a newsvendor inventory model with two replenishment decisions: the regular order and an extraordinary order. We suppose that the size of the extraordinary order depends on the behavior of the customers and it is determined as a variable fraction of the extent of shortage in the inventory. The backlogged demand rate is described by a non-increasing cosinusoidal-type function with respect to the amount of shortage. The objective is to maximize the expected total profit for the period, when the demand follows an exponential distribution. The uniqueness and existence of optimal policies are proved and, by using closed-form expressions, we determine the optimal lot size and the maximum expected profit. This work extends several newsvendor inventory models proposed in the literature.

Keywords: newsvendor model, backlogged demand rate, maximum expected profit.

AMS classification: AMS 90B05.

§1. Introduction

The newsvendor problem is a well-known stochastic inventory problem, which was initially formulated as follows. A product can be acquired only at the beginning of a selling period. The unit purchasing cost and the unit selling price are independent of the quantity of units acquired. The probability distribution of demand is known and the units remaining at the end of the period cannot be sold. The objective is to determine the optimal number of units to have stored at the start of the period to maximize the expected profit during the period. Later on, many researchers have extended this model in several ways. Thus, Gallego and Moon (1993) considered the possibility of an emergency order to provide any unsatisfied demand during the selling season with an additional charge. Khouja (1996) generalized that model allowing that, when the system is out of stock, only a fixed fraction of demand is served with delay through the extraordinary order and, therefore, the remaining fraction of demand is lost.

However, in some real inventory systems, it can be observed that the extent of the shortage determines whether the backorder will be accepted or not. In consequence, the fraction of backordered shortages is variable and depends on the unsatisfied demand. To reflect this phenomenon, Lodree (2007) proposed a non-increasing linear function of the magnitude of shortage. Recently, Lee and Lodree (2010) present two different functions to model the behavior of the customers faced with shortages.

This idea that only a fraction of the demand is served late during the period without existences was previously used in the context of inventory models with continuous review. For example, Abad (1996) considered that the fraction of backordered shortages is variable and depends on the duration of the waiting time up to the arrival of the next replenishment. He proposed two functions to model this situation. In the last few years, many OR researchers have developed different types of inventory models with partial backordering where the backlogging rate is a function dependent on the length of the waiting time. Among them, we can refer to the works of San-José et al. (2006, 2009), Dye (2007), Abad (2008), Bhunia et al. (2009), etc. Other authors have supposed that the behavior of the customers faced with a shortage depends on the lapse of time from the break in the stock (i.e., the fraction of accumulated demand depends on the net inventory level) as occurs in the model of Padmanabhan and Vrat (1990).

In this work, we consider a newsvendor problem with demand exponentially distributed and two ordering opportunities: the regular order and an extraordinary order. We suppose that the size of the extraordinary order is determined by using a non-increasing sinusoidal-type function which depends on the extent of shortage. We analyze the model, calculate the revenues and the costs related to the inventory system and present closed-form expressions to obtain the optimal lot size and the maximum expected profit. Also, a sensitivity analysis of the optimal lot size and the optimum expected profit with respect to some major parameters is carried out. Finally, we check that several newsvendor inventory models studied by other authors are particular cases of the model analyzed here.

§2. Notation and assumptions

We consider a newsvendor problem in which the demand of the product during the selling season is described by a continuous random variable X exponentially distributed with mean value μ, that is,

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}, \text{ for } x \geq 0$$

Moreover, we suppose that if the demand during the selling season x is greater than the stock size Q, then the vendor has the possibility to order a certain fraction $\beta(y)$ of the shortage y (i.e., $y = x - Q$). This extraordinary order can be ordered to the same or another manufacturer at a unit purchasing cost c_B greater than the initial unit purchasing cost c. Hence, $\omega = c_B - c > 0$ denotes the unit extra cost of the extraordinary order.

We will denote by v (greater than c) the unit selling price, and by c_H the unit effective holding cost for surplus items (which can be negative when there exists the possibility of selling them at a bargain price smaller than the unit purchasing cost, that is, $-c_H < c$). Consequently, we suppose that the total unit overstocking cost is $h = c + c_H > 0$. Moreover, we consider that each item finally not served causes a unit goodwill cost c_G in addition to the unit cost for loss of profit $v - c$ and, thus, the total unit cost of lost sales is $p = c_G + v - c > 0$. Furthermore, as in Khouja (1996), Lodree (2007) and Lee and Lodree (2010), we assume that this cost p is greater than the unit extra cost of the extraordinary order ω because, otherwise, the vendor would prefer to lose the sale rather than recover it by the extraordinary order.

Finally, we assume that the fraction $\beta(y)$ of shortage served with the extraordinary order

when the shortage is y is a non-increasing truncated cosinusoidal function, which is described by the function

$$\beta(y) = \begin{cases} \beta_o \cos\left(\frac{\pi y}{2M}\right) & \text{if } 0 \leq y \leq M \\ 0 & \text{if } y > M \end{cases}, \text{ with } M > 0 \tag{1}$$

Note that, if the shortage tends to zero, it may not satisfy all the demand and, therefore, β_o (called the extraordinary intensity) is another parameter of the system that represents the initial ratio of shortage satisfied with the extraordinary order. Also M is the maximum allowable quantity of unsatisfied demand (that is, if the shortage is greater than M, then all items are lost sales). This function generalizes the one considered in Lee and Lodree (2010), which is obtained when $\beta_o = 1$.

§3. The mathematical model

According to the previous assumptions, the objective is to maximize the expected profit. It is obvious that we firstly need to determine the total profit for the inventory system, which includes the following components: ordinary sales income, revenues due to sales of backlogged demand, initial purchasing cost, cost of the extraordinary order, effective holding cost and goodwill cost for lost sales. Of course, the ordinary sales income is $v\min(Q,x) = vx - v(x-Q)^+$ and the revenues from sales of backlogged demand is $v(x-Q)^+\beta((x-Q)^+)$. The initial purchasing cost is $cQ = cx + c(Q-x)^+ - c(x-Q)^+$ and the cost of the extraordinary order is $c_B(x-Q)^+\beta((x-Q)^+)$. Since the effective holding cost is $c_H(Q-x)^+$ and the goodwill cost for lost sales is $c_G(x-Q)^+\left[1-\beta((x-Q)^+)\right]$, we obtain that the total profit for the inventory system is

$$P(Q,x) = (v-c)x - h(Q-x)^+ - \omega(x-Q)^+\beta((x-Q)^+) - p(x-Q)^+[1-\beta((x-Q)^+)]. \tag{2}$$

In consequence, the expected profit is

$$B(Q) = (v-c)\mu - T(Q), \tag{3}$$

where

$$T(Q) = h(Q-\mu) + \int_Q^\infty \left[h + p + (\omega - p)\,\beta(x-Q)\right](x-Q)\frac{1}{\mu}e^{-\frac{x}{\mu}}dx. \tag{4}$$

Therefore, maximizing the expected profit $B(Q)$ is equivalent to minimizing the function $T(Q)$.

3.1. Solution

Substituting the backorder rate function into (4), and using the change of variable $y = x - Q$, it follows that

$$T(Q) = h(Q-\mu) + \left[h + p + (\omega - p)\,\theta\right]\mu e^{-\frac{Q}{\mu}}, \tag{5}$$

where θ is a non-negative constant independent of Q, which represents the ratio between the expected size of the extraordinary order and the expected demand. That is,

$$\theta = \frac{\int_0^\infty y\beta(y)\frac{1}{\mu}e^{-\frac{y}{\mu}}dy}{\mu} = \frac{\beta_o}{\mu^2}\int_0^M y\cos\left(\frac{\pi y}{2M}\right)e^{-\frac{y}{\mu}}dy.$$

Using the change of variable $y = \mu z$, we have

$$\theta = \beta_o \int_0^{M/\mu} ze^{-z}\cos\left(\frac{\pi\mu z}{2M}\right)dz. \tag{6}$$

Solving the above defined integral (see Appendix), we have

$$\theta = \frac{\beta_o}{1+\left(\frac{\pi\mu}{2M}\right)^2}\left[1+\left(\frac{\pi}{2}\right)e^{-M/\mu} - \frac{2-\left(\frac{4M}{\pi\mu}\right)e^{-M/\mu}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\right]. \tag{7}$$

Note that $\theta \leq \beta_o$, because $0 \leq \beta(y) \leq \beta_o$ for all $y \geq 0$. Moreover, it is easy to see that $\theta = 0$ if and only if $\beta_o = 0$ and that $\theta = 1$ if and only if $\beta_o = 1$ and $M = +\infty$.

Next, we present our main results, which determine the optimal inventory policy.

Theorem 1. *Suppose that demand is described by a random variable X with exponential distribution and expected value μ. The backorder rate function is given by (1). Then:*

1. *The function $T(Q)$ given by (5) is strictly convex on $[0,\infty)$.*
2. *The optimal lot size is*

$$Q^* = \mu\ln\left\{1+\frac{p}{h}+\frac{(\omega-p)\beta_o}{h\left[1+\left(\frac{\pi\mu}{2M}\right)^2\right]}\left[1+\left(\frac{\pi}{2}\right)e^{-M/\mu}-\frac{2-\left(\frac{4M}{\pi\mu}\right)e^{-M/\mu}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\right]\right\} \tag{8}$$

3. *The maximum expected profit is*

$$B(Q^*) = (v-c)\mu - hQ^*. \tag{9}$$

Proof. 1. Since $h+p+(\omega-p)\,\theta = h+\omega\theta+p(1-\theta) > 0$, the equation (5) shows that the function $T(Q)$ is the sum of an affine function and a strictly convex function. This is our first assertion.

2. Taking into account that the first derivative of the function $T(Q)$ is

$$T'(Q) = h - [h+p+(\omega-p)\,\theta]\,e^{-\frac{Q}{\mu}}, \tag{10}$$

we have $T'(0) = -[p+(\omega-p)\,\theta] = -[\omega\theta+p(1-\theta)] < 0$ and $\lim_{Q\to\infty}T'(Q) = h > 0$. From this, we conclude that the function $T(Q)$ attains its global minimum at

the unique solution of the equation $T'(Q) = 0$. Using the formula (10), we obtain that this solution is given by

$$Q^* = \mu \ln \left\{ 1 + \frac{p + (\omega - p)\theta}{h} \right\}. \tag{11}$$

Now substituting the value of θ defined by (7) into (11), we obtain (8).

3. Substituting (8) into (5), we can assert that $T(Q^*) = hQ^*$ and, by (3), the proof is complete.

□

3.2. Particular cases

Next, we show that several newsboy models studied by other authors can be considered as particular cases of the model developed in this paper.

1. *Basic newsboy model with demand exponentially distributed.* It is obtained from our model when $\beta_o = 0$ is considered. Now, from (6), $\theta = 0$ and, from (11), we have $Q^* = Q_o^* = \mu \ln(1 + p/h) = F^{-1}[p/(p + h)]$, where F denotes the distribution function of X. Thus, the optimal solution here shown coincides with the one given in the literature (see, for instance, Hillier and Lieberman 2001).

2. *Newsboy model with fixed partial backlogging and demand exponentially distributed (Khouja, 1996).* It is obtained from our model when we take $M \to \infty$. Therefore, we have $\beta(y) = \beta_o$. In consequence, $Q^* = Q_K^* = \mu \ln \left(1 + \frac{p+(\omega-p)\beta_o}{h}\right) = F^{-1}\left[\frac{p+(\omega-p)\beta_o}{h+p+(\omega-p)\beta_o}\right]$, which coincides with the given solution by Khouja.

3. *Newsboy model considered by Lee and Lodree (2010) with exponential demand.* This model is obtained taking $\beta_o = 1$ in the model analyzed here. In this case, $Q^* = \mu \ln \left\{ 1 + \frac{p}{h} + \frac{2(\omega-p)M^2}{h(4M^2+\pi^2\mu^2)^2} \left[2(4M^2 - \pi^2\mu^2) + \pi e^{-M/\mu}(4M^2 + 8M\mu + \pi^2\mu^2 \right] \right\}$.

§4. Sensitivity analysis

In this section, we analyze the variation of the optimal order quantity Q^* and the maximum expected profit $B(Q^*)$ with respect to some parameters of the inventory system. Since the auxiliary parameters h, ω and p depend on the initial parameters c_H, c_B, c_G, c and v, we will do the study considering these last parameters. Thus, we can rewrite (11) and (9) as

$$Q^* = \mu \ln \left\{ 1 + \frac{\theta c_B + (1 - \theta)(c_G + v) - c}{c_H + c} \right\} \tag{12}$$

and

$$B(Q^*) = (v - c)\mu - (c_H + c)Q^*. \tag{13}$$

Next, we analyze the variation of the optimal order quantity with respect to the initial parameters of the inventory model.

Theorem 2. *If the demand follows an exponential distribution with expected value μ and the backorder rate $\beta(y)$ is given by (1), then the optimal lot size Q^* verifies:*

1. *increases with c_B if $\theta > 0$ and it does not depend on c_B if $\theta = 0$;*
2. *increases with c_G if $\theta < 1$ and it does not depend on c_G if $\theta = 1$;*
3. *increases with v if $\theta < 1$ and it does not depend on v if $\theta = 1$;*
4. *decreases as c increases;*
5. *decreases as c_H increases;*
6. *increases as μ increases;*
7. *decreases as β_o increases;*
8. *decreases as M increases.*

Proof.

1. If $\theta = 0$, the optimal order quantity Q^* does not depend on c_B. However, if $\theta > 0$ then the fraction $\frac{\theta c_B + (1-\theta)(c_G+v)-c}{c_H+c}$ (consequently, also Q^*) increases as c_B increases.
2. If $\theta = 1$, the optimal order quantity Q^* does not depend on c_G. On the other hand, if $\theta < 1$ then the fraction $\frac{\theta c_B + (1-\theta)(c_G+v)-c}{c_H+c}$ increases as c_G increases and, hence, Q^* increases as the unit goodwill cost for lost sale increases.
3. This follows as in previous paragraph.
4. The ratio $\frac{\theta c_B + (1-\theta)(c_G+v)-c}{c_H+c}$ decreases as c increases because the numerator reduces and the denominator enlarges. Thus, we conclude that the optimal order quantity Q^* decreases as c increases.
5. It is clear that the ratio $\frac{\theta c_B + (1-\theta)(c_G+v)-c}{c_H+c}$ decreases as c_H increases. Therefore, from (12), Q^* decreases with the unit effective holding cost.
6. From (6), we obtain $\frac{d\theta}{d\mu} = -\frac{\pi\beta_o}{2M}\int_0^{M/\mu} z^2 e^{-z}\sin\left(\frac{\pi\mu z}{2M}\right) dz < 0$. Thus, θ decreases as μ increases. Moreover, since $c_B < c_G + v$, we see that the ratio $\frac{\theta c_B + (1-\theta)(c_G+v)-c}{c_H+c}$ increases as μ increases. Consequently, Q^* increases as the mean demand increases.
7. From (7), we see that θ increases as β_o increases. Thus, we conclude from (12) that Q^* decreases as the extraordinary intensity increases.
8. This follows as in previous paragraph.

□

The following result analyzes the sensitivity of the optimal expected profit with respect to some parameters of the inventory system.

Corollary 3. *Under the assumptions of Theorem 2, the maximum expected profit $B(Q^*)$ verifies the following sentences:*

1. *decreases as c_B increases if $\theta > 0$, and it does not depend on c_B if $\theta = 0$;*
2. *decreases as c_G increases if $\theta < 1$, and it does not depend on c_G if $\theta = 1$;*
3. *decreases as c_H increases;*
4. *decreases as c increases;*
5. *increases as v increases;*
6. *increases as β_o increases;*
7. *increases as M increases.*

Proof. It follows easily from the model formulation (see equations (2–4)), Theorem 2 and equation (13).

□

4.1. Numerical example

Next, we include a numerical example to illustrate the proposed model and its solution procedure.

Example (Taken from Lodree (2007) and adapted here to our model). Let us consider an inventory system with the assumptions assumed in this paper for which the parameters are: $c = 75$, $c_H = 20$, $c_B = 95$, $v = 115$, $c_G = 10$, $\mu = 150$, $M = 50$ and $\beta_o = 0.9$. Following the notation given in Section 2, we obtain $h = 95$, $\omega = 20$ and $p = 50$. From (7), we get $\theta = 0.019$ (that is, the expected size of the extraordinary order represents 1.9% of the expected demand). Now, applying Theorem 1, we obtain $Q^* = 62.82$ and $B(Q^*) = 31.89$.

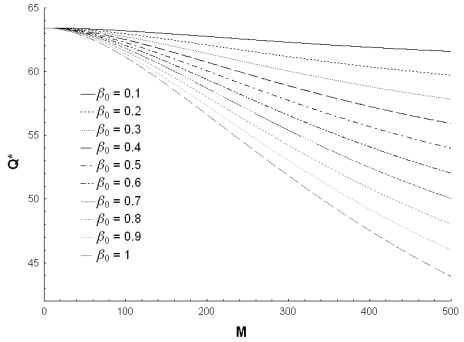


Figure 1: Q^* as function of M and β_o

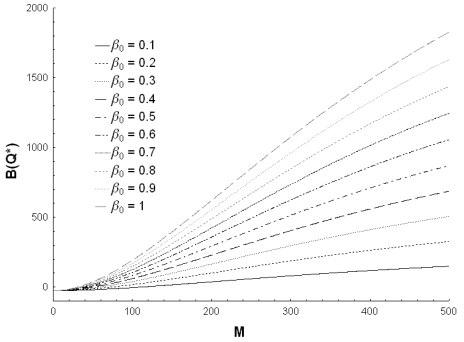


Figure 2: $B\left(Q^*\right)$ as function of M and β_o

In Fig.1 we plot the variation of the optimal lot size Q^* as function of the parameters β_o and M. According to Theorem 2, the figure shows that the optimal order quantity decreases if M or β_o are increasing.

Fig. 2 shows that, for a fixed β_o, the maximum expected profit increases with M. In the same way, for a fixed M, if β_o is increasing, we have a higher optimal expected profit, as is asserted in Corollary 3.

§5. Conclusions

Inventory systems in which the fraction of backlogged demand depends on the unsatisfied demand are based on the realistic observation of the customers' behavior faced with a shortage. We analyze a newsvendor inventory model with demand exponentially distributed and two orders: the regular lot size and an extraordinary order. We consider that the size of the extraordinary order depends on the behavior of the customers and it is described by a non-increasing sinusoidal-type function which depends on the extent of shortage. After proving the uniqueness and existence of optimal decisions, we determine the optimal order quantity and the maximum expected profit using closed-form expressions. Also we develop a sensitivity analysis of the optimal policy and the maximum expected profit with respect to the parameters of the inventory system. Thus, for instance, we show that the optimal lot size increases as the unit selling price, or the unit goodwill cost for lost sale, or the unit cost for the extraordinary order increases. Nevertheless, the optimal order decreases if the unit effective holding cost, or the unit purchasing cost, or the extraordinary intensity increases.

The model can be extended in several ways. For instance, we could consider other probability distributions for the customers' demand. Also, we could assume other backorder rate functions or to consider multiple selling periods.

Acknowledgements

This work is partly supported by Spanish Ministry of Science and Innovation through the research project MTM2010-18591.

Appendix

Let us consider the defined integral $\int_0^{M/\mu} ze^{-z}\cos\left(\frac{\pi\mu z}{2M}\right)dz$. Taking into account the following two integrals

$$\int e^{-z}\cos\left(\frac{\pi\mu z}{2M}\right)dz = \frac{\frac{2M}{\pi\mu}e^{-z}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\left[\sin\left(\frac{\pi\mu z}{2M}\right)-\frac{2M}{\pi\mu}\cos\left(\frac{\pi\mu z}{2M}\right)\right]$$

$$\int e^{-z}\sin\left(\frac{\pi\mu z}{2M}\right)dz = \frac{-\frac{2M}{\pi\mu}e^{-z}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\left[\cos\left(\frac{\pi\mu z}{2M}\right)+\frac{2M}{\pi\mu}\sin\left(\frac{\pi\mu z}{2M}\right)\right]$$

and by using the integration by parts method, we have

$$\begin{aligned}
\int_0^{M/\mu} ze^{-z}\cos\left(\frac{\pi\mu z}{2M}\right)dz &= \frac{1}{1+\left(\frac{2M}{\pi\mu}\right)^2}\left[\frac{2M^2}{\pi\mu^2}e^{-M/\mu}-\frac{2M}{\pi\mu}\int_0^{M/\mu}e^{-z}\sin\left(\frac{\pi\mu z}{2M}\right)dz\right.\\
&\qquad\left.+\left(\frac{2M}{\pi\mu}\right)^2\int_0^{M/\mu}e^{-z}\cos\left(\frac{\pi\mu z}{2M}\right)dz\right]\\
&= \frac{\frac{2M^2}{\pi\mu^2}e^{-M/\mu}-\left(\frac{2M}{\pi\mu}\right)^2\left[\frac{1-\left(\frac{2M}{\pi\mu}\right)^2-\left(\frac{4M}{\pi\mu}\right)e^{-M/\mu}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\right]}{1+\left(\frac{2M}{\pi\mu}\right)^2}\\
&= \frac{1}{1+\left(\frac{\pi\mu}{2M}\right)^2}\left[1+\left(\frac{\pi}{2}\right)e^{-M/\mu}-\frac{2-\left(\frac{4M}{\pi\mu}\right)e^{-M/\mu}}{1+\left(\frac{2M}{\pi\mu}\right)^2}\right]
\end{aligned}$$

References

[1] P.L. ABAD. Optimal pricing and lot-sizing under conditions of perishability and partial backordering. *Management Science 42* (1996), 1093–1104.

[2] P.L. ABAD. Optimal price and order size under partial backordering incorporating shortage, backorder and lost sale costs. *International Journal of Production Economics 114* (2008), 179–186.

[3] A.K. BHUNIA, S. KUNDU, T. SANNIGRAHI T. AND S.K. GOYAL. An application of tournament genetic algorithm in a marketing oriented economic production lot-size model for deteriorating items. *International Journal of Production Economics 119* (2009), 112–121.

[4] C.Y. DYE. Joint pricing and ordering policy for a deteriorating inventory with partial backlogging. *Omega 35* (2007), 184–189.

[5] G. GALLEGO AND I. MOON. The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society 44* (1993), 825–834.

[6] F.S. HILLIER AND G.J. LIEBERMAN. *Introduction to Operations Research, seventh ed.* McGraw-Hill, Boston, 2001.

[7] M.A. KHOUJA. A Note on the newsboy problem with an emergency supply option. *Journal of the Operational Research Society 47* (1996), 1530–1534.

[8] H. LEE AND E.J. LODREE. Modeling customer impatience in a newsboy problem with time-sensitive shortages. *European Journal of Operational Research 205* (2010), 595–603.

[9] E.J. LODREE. Advanced supply chain planning with mixtures of backorders, lost sales, and lost contract. *European Journal of Operational Research 181* (2007), 168–183.

[10] G. PADMANABHAN AND P. VRAT. Inventory model with a mixture of backorders and lost sales. *International Journal of Systems Sciences 21* (1990), 1721–1726.

[11] L.A. SAN JOSÉ, J. SICILIA AND J. GARCÍA-LAGUNA. Analysis of an inventory system with exponential partial backordering. *International Journal of Production Economics 100* (2006), 76–86.

[12] L.A. SAN-JOSÉ, J. GARCÍA-LAGUNA AND J. SICILIA. An economic order quantity model with partial backlogging under general backorder cost function. *Top 17* (2009), 366–384.

Valentín Pando
Departamento de Estadística e Investigación Operativa
Universidad de Valladolid
vpando@eio.uva.es

Luis A. San-José
Departamento de Matemática Aplicada
Universidad de Valladolid
augusto@mat.uva.es

Juan García-Laguna
Departamento de Estadística e Investigación Operativa
Universidad de Valladolid
laguna@eio.uva.es

Joaquín Sicilia
Departamento de Estadística, Investigación Operativa y Computación
Universidad de La Laguna
jsicilia@ull.es

Monografías Matemáticas García de Galdeano **38**, 117–127 (2013)

SOME BASIC STATISTICS OF GENERAL RENEWAL PROCESSES

Javier Villarroel

Abstract. We consider a random processes whose evolution in time results from the combined effect of a constant drift and the occurrence of random jumps. The jump part is modelled by a classical compound renewal process, namely a compound Poisson process generalized to have arbitrary i.i.d. waiting times. Such models are of overriding interest in insurance and ruin theory. The problem of determining the exit time from a given interval is reduced to a renewal integral equation. We consider in particular the case of Erlang waiting times.

Keywords: mean exit times, renewal stochastic processes.

AMS classification: 91B30, 60K15, 60J75.

§1. introduction

Stochastic jump models have a venerable story of paramount importance in probability and risk theory [1, 2, 3], and as such have been used to model statistics of a multitude of random phenomena (see also[4, 5]). To list a few early examples we note applications to earthquake modelling (e.g., [6, 7]), rainfall description [8, 9] and the statistics of flare activity in stars [10]. Applications to describe changes of stock markets due to unexpected catastrophes were first noted in the seminal work of Merton [11] where it is assumed that inter-catastrophe times are exponentially distributed and independent of the magnitude of the catastrophe, i.e., that catastrophes are driven by a compound Poisson process (CPP). Jump processes have been also used widely in actuarial and financial studies, [12, 13, 14].

The paradigmatic and simplest model which underlies all these situations is the classical Poisson process. Compound renewal processes constitute a natural generalization of the latter. They are obtained considering two sequences of positive random variables $\{\tau_n\}_{n=1\ldots\infty}$ and $\{J_n\}_{n=1\ldots,\infty}$ defined on a certain probability space which satisfy the assumptions:

(i) τ_n are independent and identically distributed random variables (i.i.d.r.v.) with probability density (PDF) and cumulative distribution function $\psi(t)$ and $\Psi(t) = \int_0^t \psi(t')dt'$;

(ii) J_n is a sequence of i.i.d.r.v. with common PDF $h(\cdot)$;

(iii) J_m is independent of τ_n for any n, m.

The "arrival times" are defined by $t_0 = 0,\ t_n \equiv \tau_1 + \cdots + \tau_n, n \geq 1$ while we call $\tau_n \equiv t_n - t_{n-1} > 0$ the "waiting times". In terms of these variables the renewal process is the increasing function $t \mapsto N_t$ defined by $N_t = n$ for t on the interval $[t_n, t_{n+1})$, i.e.,

$$N_t = \sum_{n=0}^{\infty} n\mathbf{1}_{\{t_n \leq t < t_{n+1}\}} \tag{1}$$

while the more general object

$$S_t = \sum_{n=0}^{\infty} (J_1 + \cdots + J_n) \mathbf{1}_{\{t_n \leq t < t_{n+1}\}} \equiv \sum_{n=0}^{N_t} J_n \tag{2}$$

is called a compound-renewal process (see figure 1). Note that S_t (respectively N_t) takes a constant value $S_t = J_1 + \cdots + J_n$ (respectively n) on the interval $[t_n, t_{n+1})$ while both $S_t = N_t = 0$ if $t < t_1$. Further S_t has right-continuous and piece wise constant sample paths $t \mapsto S_t$, with jumps discontinuities at t_n at which S_t has a jump $S_{t_n^+} - S_{t_n^-} = J_n$ (respectively, $N_{t_n^+} - N_{t_n^-} = 1$).

Poisson process N_t corresponds to having holding times τ_n exponentially distributed $\tau_n \sim \mathcal{E}(\lambda)$ for some $\lambda > 0$; then N_t has Poisson distribution $N_t \sim \mathcal{P}(\lambda t)$. Further S_t is termed the compound Poisson process (CPP). In this case both N_t and S_t are Markovian, independent-increments processes with right-continuous sample paths, namely Levy processes.

Here we are interested in more general random processes X_t whose evolution in time can be thought of as the result of the combined effect of a constant drift and the occurrence of random jumps, i.e., compound renewal processes with drift. Thus we can write

$$X_t = vt - \sum_{n=0}^{N_t} J_n \equiv vt - S_t \tag{3}$$

where $v \geq 0$ is a constant (the drift) and J_n is jump. J_n represents the sudden variation of a statistical observable (amount of rainfall in a certain shower etc) and N_t and S_t are defined above, cf. eqs. (1) and (2)

This further addition of the drift term is a natural and significant incorporation. The resulting process plays a fundamental role in, say, actuarial studies. Here N_t and S_t represent, respectively, the number of claims in $[0, t]$ and the aggregate claims arriving at a non-life insurance company over the time period $[0, t]$. Finally, X_t- the surplus process- is a prototype model in risk management to describe the dynamics of the cahsflow at an insurance company under the assumption that premiums are received at a constant rate $v > 0$ and that the company incurs in losses $J_n > 0$ from claims reported at times $t_n, n = 0, \ldots \infty$. It was introduced by Cramer-Lundberg and later generalized by Sparre Anderson to have arbitrary i.i.d. waiting times, see [12, 13, 14]. See also [15, 16, 17]. More recently, it has been shown that this process also rules the rate of energy dissipation in nonlinear optical fibers [18, 19]. In both scenarios a problem that arises naturally is that of determining the first exit time off a given interval [20].

The consideration of these general models is motivated by the believe that the exponential holding-time assumption underlying the Cramer-Lundberg model may, in many settings, be inadequate to describe the actuarial situation; the Sparre Anderson model gives more flexibility to fit adequately both waiting times and sample paths. This liberty could be of interest to capture various stylized features observed in the data like, say, "heavy tails" in the waiting times distribution or a renewal function $m(t) \equiv \mathbb{E}\big(N_t\big)$ which departs from linearity. Properties of the compound Poisson surplus process, though already considered by Cramer, still remain an important topic of research.

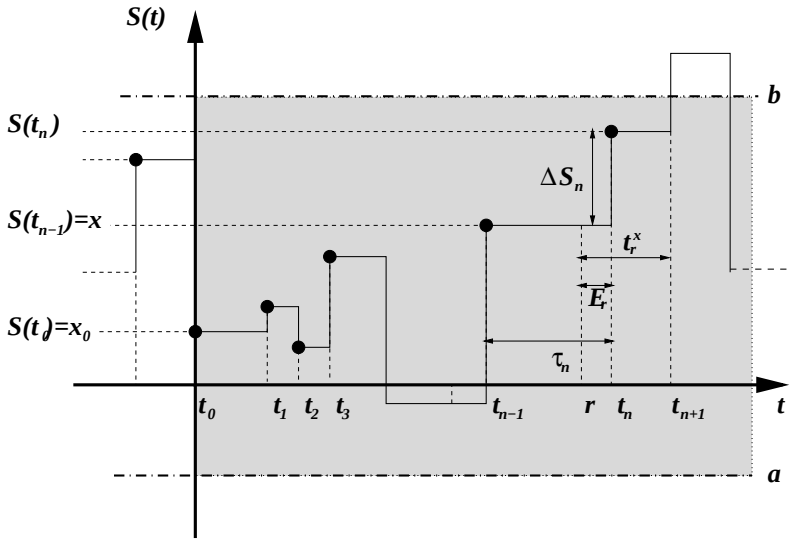


Figure 1: A typical sample path $t \mapsto S_t$ of a compound Poisson process with arrival times $\{t_n\}_{n=0,\dots\infty}$ and waiting times τ_n. $\mathbf{t}_r^x$ is the time to exit the interval (a, b) starting from x at time r where $t = r$ represents the present instant.

It turns out that the combined effects of having general waiting times and, in addition, the incorporation of a drift, render quite difficult the study of the different statistical properties of the model. Further, generically N_t and S_t are not Markov processes (cf. Prop. 1 below). Hence results derived at jump times do not extend to generic present. Hence the issue on how these results must be corrected when the *present instant is not a time at which a claim occurs* follows in a natural way. We find that the solution to these problems is contingent on *the available information*. However, due to the ensuing difficulty these more general situations will be skipped (see nevertheless some comments on this regard at the end of the paper).

§2. Mean exit times

Let X_t be a renewal-point process with drift, $r \geqslant 0$ the actual time and suppose that $X_r = x$ where $x, 0 < x < \xi$, is the actual position and ξ a reference level. A fundamental statistics of the problem is the first exit time of the process from $(0, \xi)$ (see Fig. (1)). We pose the problem of evaluating such time given that the present is one of the jump times. In insurance this corresponds to the problem of evaluating the mean time for the insurance company capital X_t to reach a given level ξ or get bankrupt before. We assume that $v > 0$ and the jumps J_n are positive random variables so that $vJ_n > 0$. With this choice the effect of the drift is to increase the process towards ξ while jumps have an opposite effect. Without proof we note the following:

Proposition 1. *The process X_t is only pseudo-Markovian* . Concretely,

(i)The associated "skeleton" process $\{Y_n \equiv X_{t_n}\}_{n=0\ldots,\infty}$ is a *discrete time* Markov chain for all choices of jump density $h(\cdot)$.

(ii) The continuous-time process X_t is Markovian iff the waiting time distribution is exponential $\psi(t) = \lambda e^{-\lambda t}$.

As a consequence under the exponential assumption the strong Markov property holds and *results derived at a jump time extend to arbitrary present*. However no such inference is possible with more general waiting-times distribution since then Markoviannes is lost.

2.1. Mean time renewal equation

We now derive the integral equation that the mean escape time satisfies. Here we consider only the mean time when the present is a jump-time $r \equiv t_n$. Given $X_{t_n} = x$ let $t_n + \mathbf{t}^x_{t_n}$ be the first time after t_n at which the process exits $(0,\xi)$ where $0 < x < \xi$, namely $t_n + \mathbf{t}^x_{t_n} = \inf_t\{t \geq t_n : X_t \notin (0,\xi)\}$. We note that $\{\mathbf{t}^x_{t_n}\}$ is a sequence of i.i.d.r.v. whose distribution depends only on x but does not depend on either n, t_n or the "history" of the process $\sigma(X_s, s \leq t_n)$. Hence denote by $\mathbb{M}(x) = \mathbb{E}\big(\mathbf{t}^x_{t_n}\big)$ the mean of $\mathbf{t}^x_{t_n}$ and $\mathbb{E}[\cdot]$ the expectation operator.

We aim to determine the mean of $\mathbf{t}^x_{t_n}$. Recalling the choice of signs, cf. (3), it follows that exit through the upper barrier ξ can *only* happen through the drift effect vt while, by contrast, escape below the lower end 0 will stem from the jump term. The key fact to realize is that after t_n three possibilities arise: if $\tau_{n+1} \geq \varrho \equiv \frac{\xi - x}{v}$ (we recall that $\tau_{n+1} \equiv t_{n+1} - t_n$) then the drift pushes X_t to reach the level ξ at time $t_n + \varrho$. Otherwise, and if a jump of magnitude $J \equiv J_{n+1}$ occurs at t_{n+1} such that $x + v\tau_{n+1} - J \leq 0$ then $X_{t_{n+1}}$ goes below zero and exits the interval $(0,\xi)$. Finally, if this jump satisfies $x + v\tau_{n+1} - J > 0$ the process remains within $(0,\xi)$ and starts afresh with a "surplus" $X_{t_{n+1}} = x + v\tau_{n+1} - J$ and the time remaining to exit will be $\mathbf{t}^{X_{t_{n+1}}}_{t_{n+1}}$. Thus on $\{X_r = x\}$ is

$$t_n + \mathbf{t}^x_{t_n} = (t_n + \varrho)\mathbf{1}_{\{\tau_{n+1} \geq \varrho\}} + (t_n + \tau_{n+1})\mathbf{1}_{\{\tau_{n+1} < \varrho\}}\mathbf{1}_B +$$

$$(t_n + \tau_{n+1} + \mathbf{t}^{X_{t_{n+1}}}_{t_{n+1}})\mathbf{1}_{\{\tau_{n+1} < \varrho\}}\mathbf{1}_{B^c} \tag{4}$$

where we define the event $B \equiv \{x + v\tau_{n+1} - J \leq 0\}$. Rearranging several terms we see that $\mathbf{t}^x_{t_n}$ must satisfy the functional equation

$$\mathbf{t}^x_{t_n} = \varrho\mathbf{1}_{\{\tau_{n+1} \geq \varrho\}} + \tau_{n+1}\mathbf{1}_{\{\tau_{n+1} < \varrho\}} + \mathbf{t}^{X_{t_{n+1}}}_{t_{n+1}}\mathbf{1}_{\{\tau_{n+1} < \varrho\}}\mathbf{1}_{B^c} \tag{5}$$

Next, taking conditional expectations one has

$$\mathbb{E}\Big(\varrho\mathbf{1}_{\{\tau_{n+1} \geq \varrho\}} + \tau_{n+1}\mathbf{1}_{\{\tau_{n+1} < \varrho\}}\Big| X_{t_n} = x\Big) = \varrho\Big(1 - \Psi\big(\varrho\big)\Big) + \int_0^\varrho l d_l \Psi(l)$$

$$= \int_0^\varrho \Big(1 - \Psi\big(l\big)\Big) dl$$

Further, we use the well known tower property of conditional expectation and also that τ_{n+1} and $x + v\tau_{n+1} - J$ are $\sigma(\tau_{n+1}, J)$-measurable. Then, if $J = J_{n+1}$ we can evaluate the expectation of the last term in Eq. (5) as

$$\begin{aligned}
&\mathbb{E}\Big(\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \mathbf{1}_{\{\tau_{n+1}<\varrho\}} \mathbf{1}_{B^c} \Big| X_{t_n} = x\Big) = \\
&\mathbb{E}\Big(\mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \mathbf{1}_{\{\tau_{n+1}<\varrho\}} \mathbf{1}_{B^c} \Big| X_{t_n} = x, \tau_{n+1}, J\Big] \Big| X_{t_n} = x\Big) = \\
&\mathbb{E}\Big(\mathbf{1}_{\{\tau_{n+1}<\varrho\}} \mathbf{1}_{B^c} \cdot \mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \Big| X_{t_n} = x, \tau_{n+1}, J\Big] \Big| X_{t_n} = x\Big)
\end{aligned}$$

$$\int \mathbf{1}_{\{l<\varrho\}} \mathbf{1}_{\{x+vl>y\}} \mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \Big| \tau_{n+1} = l, J = y, X_{t_n} = x\Big] \mathbb{P}\big(\tau_{n+1} \in dl, J \in dy | X_{t_n} = x\big)$$

To proceed further notice that since X_{t_n} is $\sigma(\tau_1, \dots \tau_n, J_1, \dots J_n)$-measurable, the model assumptions imply that τ_{n+1} and J_{n+1} are independent of X_{t_n}:

$$\mathbb{P}\big(\tau_{n+1} \in dl, J \in dy | X_{t_n} = x\big) = \mathbb{P}\big(\tau_{n+1} \in dl\big)\mathbb{P}\big(J \in dy\big)$$

Further, by the pseudo-Markov property , viz proposition 1, and those derived for the sequence $\mathbf{t}_{t_n}^x$ we have

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \Big| \tau_{n+1} = l, J = y, X_{t_n} = x\Big] = \\
&\mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \Big| \tau_{n+1} = l, J = y, X_{t_n} = x, X_{t_{n+1}} = x + vl - y\Big] = \\
&\mathbb{E}\Big[\mathbf{t}_{t_{n+1}}^{x+vl-y} \Big| X_{t_{n+1}} = x + vl - y\Big] = \mathbb{E}\Big[\mathbf{t}_0^{x+vl-y}\Big] = \mathbb{M}(x + vl - y)
\end{aligned}$$

Thus

$$\mathbb{E}\Big(\mathbf{t}_{t_{n+1}}^{X_{t_{n+1}}} \mathbf{1}_{\{\tau_{n+1}<\varrho\}} \mathbf{1}_{B^c} \Big| X_{t_n} = x\Big) = \int_0^{\varrho} \mathbb{P}\big(\tau_{n+1} \in dl\big) \int_0^{x+vl} \mathbb{P}\big(J \in dy\big)\mathbb{M}(x+vl-y) =$$

$$\int_0^{\varrho} \psi(l)dl \int_0^{x+vl} h(y)dy\mathbb{M}(x + vl - y) = \frac{1}{v}\int_x^{\xi} \psi\Big(\frac{l-x}{v}\Big)dl \int_0^{l} \mathbb{M}(l - y)h(y)dy$$

(letting $l \mapsto l' = x + vl$ and then dropping primes).

Collecting all these results it follows that $\mathbb{M}(u)$ satisfies the *linear integral equation*

$$\mathbb{M}(x) = \int_0^{\varrho} \Big(1 - \Psi\big(l\big)\Big)dl + \frac{1}{v}\int_x^{\xi} \psi\Big(\frac{l-x}{v}\Big)dl \int_0^{l} \mathbb{M}(l - y)h(y)dy \tag{6}$$

Thus the mean time to exit $(0, \xi)$ satisfies Eq. (6). Unfortunately the latter does not appear to be solvable in closed form. However, we show in this section that there exists a sub-class of densities for which such a solution is possible.

Remark 1. For notational convenience we have assumed that $\Psi(t) \equiv \mathbb{P}(\tau_1 \le t)$, $H(u) \equiv \mathbb{P}(J_1 \le u)$ have densities ψ and h. However the previous analysis carries over to a general case and the mean time to exit $(0, \xi)$ is found to solve the integral equation

$$\mathbb{M}(x)=\int_0^{\varrho}\Big(1-\Psi\big(l\big)\Big)dl+\frac{1}{v}\int_x^{\xi}d_l\Psi\Big(\frac{l-x}{v}\Big)\int_0^{l}\mathbb{M}(l-y)d_yH(y) \tag{7}$$

Remark 2. If $v=0$ this equation simplifies to

$$\mathbb{M}(x)=\mu+\int_0^{x}\mathbb{M}(x-y)h(y)dy \tag{8}$$

which is a classical renewal equation, [4, 5]. Here $\mu\equiv\mathbb{E}(\tau_n)=\int_0^{\infty}\Big(1-\Psi\big(l\big)\Big)dl$.

§3. Example: Exponential and Erlang cases

Here we study the mean exit time for a class of waiting time densities for which it is possible to solve the linear integral equation (6).

3.1. Exponential case: $\psi(t)=\lambda e^{-\lambda t}$

Then Eq. (6):

$$\mathbb{M}(x)=\Big(1-e^{-\lambda\varrho}\Big)/\lambda+\frac{1}{v}\int_x^{\xi}\lambda e^{-\lambda\frac{l-x}{v}}dl\int_0^{l}\mathbb{M}(l-y)h(y)dy \tag{9}$$

does not appear to be solvable in a direct way. However, by direct derivation on Eq. (9) we find that $\mathbb{M}(x)$ satisfies also the simpler equation

$$\Big(-v\partial_x+\lambda\Big)\mathbb{M}(x)=1+\lambda\int_0^{x}\mathbb{M}(x-z)h(z)dz,\ 0\leq x<\infty \tag{10}$$

along with the boundary condition $\mathbb{M}(x=\xi)=0$-which follows also from Eq. (6).

Note how the right hand side of this last equation has simplified to a renewal type term; concretely, it is of convolution type and can be solved by Laplace transformation for general choice of jump-density h. To be specific we shall consider here the case when $h(y)=\gamma e^{-\gamma y}$ is also exponential although the reasoning carries over to a general density h. We next introduce the Laplace transforms of h and ψ as the functions of the real variable $s\in\mathbb{R}^{+}$:

$$\hat{\psi}(s)\equiv\int_0^{\infty}e^{-st}\psi(t)dt=\frac{\lambda}{\lambda+s},\ \hat{h}(s)\equiv\int_0^{\infty}e^{-sy}h(y)dy=\frac{\gamma}{\gamma+s}$$

Similarly let $\hat{\mathbb{M}}(s)\equiv\int_0^{\infty}e^{-sx}\mathbb{M}(x)dx$ be the Laplace transform of the unknown function $M(x)$. We multiply Eq. (10) by e^{-sx} and integrate on x to find

$$\int_0^{\infty}e^{-sx}\Big(-v\partial_x+\lambda\Big)\mathbb{M}(x)dx=\big(\lambda-vs\big)\hat{\mathbb{M}}(s)-v\mathbb{M}_0 \tag{11}$$

$$\int_0^{\infty}e^{-sx}dx=1/s,\ \int_0^{\infty}e^{-sx}dx\lambda\int_0^{x}\mathbb{M}(x-z)h(z)dz=\lambda\hat{h}(s)\hat{\mathbb{M}}(s) \tag{12}$$

where $\mathbb{M}_0 \equiv \mathbb{M}(x=0)$ is at this stage unknown. We have used well known properties of the Laplace transform; concretely using partial integration one finds

$$\int_0^\infty e^{-sx}\mathbb{M}^{(n)}(x)dx = s^n\hat{\mathbb{M}}(s) - \Big(s^{n-1}\mathbb{M}(0) + \cdots + \mathbb{M}^{(n-1)}(0)\Big) \tag{13}$$

where $\mathbb{M}^{(j)} \equiv \partial^j\mathbb{M}$. Similarly Eq. (12) follows by interchange of integrals.

Substituting this into (10) we find that $\hat{\mathbb{M}}(s)$ must satisfy

$$\Big(\lambda - vs - \lambda\hat{h}(s)\Big)\hat{\mathbb{M}}(s) = 1/s - v\mathbb{M}_0$$

and hence $\hat{\mathbb{M}}(s)$ is given by

$$\hat{\mathbb{M}}(s) = \frac{1/s - v\mathbb{M}_0}{-vs\big(s-\varepsilon\big)}(s+\gamma) \tag{14}$$

Then $\mathbb{M}(x)$ can be recovered by the Laplace inversion formula as

$$\mathbb{M}(x) = \int_{c-i\infty}^{c+i\infty} e^{sx}\frac{1/s - v\mathbb{M}_0}{-vs\big(s-\varepsilon\big)}(s+\gamma)ds,\ c>0 \tag{15}$$

Here $\varepsilon = \frac{\lambda}{v} - \gamma$ is the so called *loading factor*, and $c>0$ is arbitrary.

To evaluate the integral (15) we take c and R to be fixed given numbers where $c \to 0^+$ and $R \to \infty$. We next complexify the variable s and consider a closed contour $C = C_1 \cup C_2$, say, on the complex s-plane consisting of (i) the line that runs from $c-iR$ to $c+iR$, (ii) a large half-circle on the left-half plane joining $c+iR$ with $c-iR$. On C_1 we can write $s = c + is_I$ where $s_I \in [-R, R]$. By contrast on C_2 we have the parametrization $s = Re^{i\varphi}, \pi/2 \le \varphi \le 3/2\pi, xcos\varphi < 0$ and

$$|e^{sx}| = |e^{xRcos\varphi}| \underset{R\to\infty}{\to} 0$$

Hence we have $\int_C = \int_{C_1} + \int_{C_2}$. In addition as R goes to infinity $\int_{C_1} =$ (15), while $\int_{C_2} \to 0$ and

$$\mathbb{M}(x) = \lim_{R\to\infty}\int_{C_1} = \lim_{R\to\infty}\int_C e^{sx}\frac{1/s - v\mathbb{M}_0}{-vs\big(s-\varepsilon\big)}(s+\gamma)ds \tag{16}$$

The integrand is a meromorphic function of the complex variable s having a (double) pole $s=0$ and a single one $s=\varepsilon$. By Cauchy theorem this integral is the sum of the residues at the poles. It follows that

$$\mathbb{M}(x) = \Big((\gamma+\varepsilon)(1-e^{\varepsilon x}) + \gamma\varepsilon x\Big)/(v\varepsilon^2) + \frac{\mathbb{M}_0}{\varepsilon}\Big((\varepsilon+\gamma)e^{\varepsilon x} - \gamma\Big) \tag{17}$$

Note that at this stage $\mathbb{M}(x)$ depends on a *free constant* $\mathbb{M}_0 \equiv \mathbb{M}(x=0)$. While apparently $\mathbb{M}_0$ should follows demanding consistency: $\mathbb{M}(x=0) = \mathbb{M}_0$ it turns out that this relation is identically satisfied. Actually $\mathbb{M}_0$ follows requiring instead the "missing" boundary condition $\mathbb{M}(x=\xi) = 0$ to hold. Solving and substituting we finally find that, in terms

of the distance to the boundary $\tilde{x} = x - \xi$, the mean time to exit $(0, \xi)$ having started at $x, 0 < x < \xi$ is

$$\mathbb{M}(x) = \frac{1}{\varepsilon v}\Big[\frac{\gamma + \varepsilon + \gamma^2 b e^{-\varepsilon\xi} - (\gamma+\varepsilon)(1+\gamma\xi)e^{\varepsilon(x-\xi)}}{\gamma + \varepsilon - \gamma e^{-\varepsilon\xi}} + \gamma x\Big]$$

$$= \frac{\gamma\tilde{x}}{\varepsilon v} + \frac{(1+\gamma\xi)}{\varepsilon v\Big(1 - \frac{\gamma e^{-\varepsilon\xi}}{\gamma+\varepsilon}\Big)}\Big[1 - e^{\varepsilon\tilde{x}}\Big] \tag{18}$$

In the case when there is no drift: $v = 0$ this expression simplifies drastically. Letting $v \to 0$ (or $\varepsilon \to \infty$) we find

$$\mathbb{M}(x) = \frac{(1+\gamma x)}{\lambda} \equiv \mathbb{E}(\tau_{n+1})\Big(1 + \frac{x}{\mathbb{E}(J_{n+1})}\Big) \tag{19}$$

The result is easy to understand since in this case obviously the process never increases; hence it can only scape $(0, \xi)$ through the lower boundary due to the jumps and not before the fist one occurs. It follows that $\mathbb{M}(x)$ can only depend on x but not on the value ξ, that $\mathbf{t}^x_{t_n} \geq \tau_{n+1}$ a.s. $\mathbb{P}$ and hence $\mathbb{M}(x) \geq \mathbb{E}(\tau_{n+1}) = 1/\lambda$.

Returning to a general case $v > 0$, notice that the mean time to reach the level 0 having started from a level x can be recovered from Eq. (18) by letting $\xi \to \infty$. This classical result is of paramount importance in risk theory, cf. [5], as it gives the mean bankruptcy time. It will be finite iff the loading factor is positive. It reads

$$\mathbb{M}_\infty(x) = \frac{1}{\varepsilon v}\Big[1 + \gamma x\Big],\ \varepsilon > 0 \text{ and } \mathbb{M}_\infty(x) = \infty \text{ if } \varepsilon \leq 0 \tag{20}$$

3.2. Erlang waiting times

A second interesting case is obtained when waiting times have Erlang distribution $\tau_n \sim \mathcal{E}r(\lambda, N)$ with $N = 1, 2 \ldots$. We recall that the Erlang distribution is obtained from the Gamma distribution when the shape parameter is an integer N and hence has density

$$\psi(t) = \lambda\big(\lambda t\big)^{N-1}\frac{e^{-\lambda t}}{(N-1)!} \tag{21}$$

With $N = 1$ we recover the exponential distribution. Here we consider the natural case $N = 2$. In this case X_t is not Markovian. Nevertheless the mean scape time solves Eq. (6).

In this case by operating with the operator $\Big(-v\partial_x + \lambda\Big)^2$ Eq. (6) simplifies to

$$\Big(-v\partial_x + \lambda\Big)^2\mathbb{M}(x) = 2\lambda + \lambda^2\int_0^x \mathbb{M}(x-z)h(z)dz,\ 0 \leq x < \infty \tag{22}$$

Note how again this equation is of convolution type and can be solved by Laplace transformation. The term $\int_0^\infty e^{-sx}\Big(-v\partial_x + \lambda\Big)^2\mathbb{M}(x)dx$ will introduce two free constants $\mathbb{M}_0, \mathbb{M}_0'$ that need to be fixed (see Eq. (13)). Eq. (6) also shows that the mean exit time to exit $(0, \xi)$

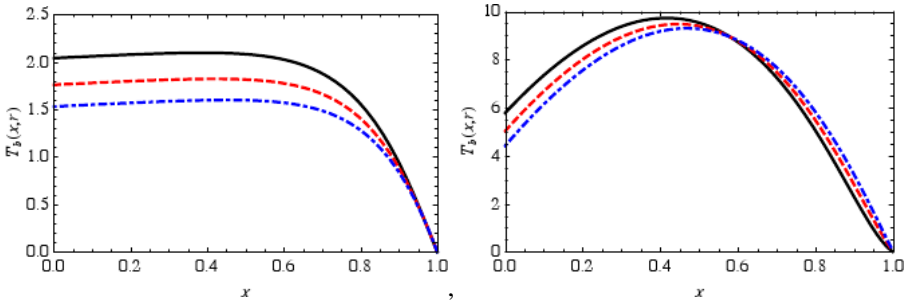


Figure 2: Different plots of $T_\xi(x,r)$ for different values of r: $r = 0$, i.e. $\mathbb{M}_\xi(x)$, solid (black) line, $r = 0.4$, dashed (red) line and $r = 10.0$, dot-dashed (blue) line. The rest of the parameters were chosen as follows, $\xi = 1.0$, $v = 0.1$, $\lambda = 1.0$ and (a) $\gamma = 0.1$; (b) $\gamma = 4.0$. In the upper panel we observe that this function is not decreasing with x and has a maximum in the interior of the interval. In the lower panel we observe a cross-over phenomenon reflecting the fact that $\mathbb{M}_\xi(x)$ needs not to be greater than $T_\xi(x,r)$.

must satisfy the boundary conditions $\mathbb{M}(\xi) = 0, \quad \mathbb{M}'(\xi) = -1/v$ which will pin down a particular solution. The solution depends in both x and ξ. Hence, let $\mathbb{M}_\xi(x)$ be such solution. To be specific we consider here the case when $h(y) = \gamma e^{-\gamma y}$ is also exponential. In this case the solution can be found after tedious calculations. The resulting expression is cumbersome and reads

$$\mathbb{M}_\xi(x) = \frac{1}{(\lambda - 2\gamma v)\Xi}\Bigg\{2\lambda qv(1+\gamma x)$$

$$+e^{-p\xi}\Bigg[qv\Big(\gamma v\big(4\gamma(\xi-x)-1\big)\cosh q\xi - 2\lambda(1+\gamma\xi)e^{px}\cosh q(\xi-x) + \gamma v e^{-p(\xi-x)}\cosh qx\Big)$$

$$+\gamma v\big(2(\lambda+\gamma v)\gamma(\xi-x) - pv\big)\sinh q\xi - \lambda\big(\lambda - \gamma v$$

$$+2pv\gamma\xi\big)e^{px}\sinh q(\xi-x) + p\gamma v^2 e^{-p(\xi-x)}\sinh qx\Bigg]\Bigg\}, \tag{23}$$

where

$$p = \frac{\lambda}{v} - \frac{\gamma}{2} \quad q = \sqrt{\frac{\lambda\gamma}{v} + \frac{\gamma^2}{4}},\ \Xi = \lambda qv - \gamma v e^{-p\xi}\Big[2qv\cosh q\xi - (\lambda+\gamma v)\sinh q\xi\Big].$$

Since the process X_t is not Markovian the exit time after jump instants, $\mathbf{t}_0^x$, say, is different to that corresponding to the "present" being a general time r. We have used a numerical simulation to recover $T(x,r)$- the mean exit time when the present r is not necessarily a jump instant. In figure 2 we plot both $T(x,r)$ and $\mathbb{M}(x) \equiv \mathbb{E}\mathbf{t}_0^x$, cf. eq. (23), for different value of parameters. Note also that one must have $T(x, r=0) = \mathbb{M}(x)$.

Finally, $\mathbb{M}_\infty(x)$ or mean time to exit $(0,\infty)$ is recovered as follows:

$$\mathbb{M}_\infty(x) = \lim_{\xi\to\infty}\mathbb{M}_\xi(x)$$

which exists only if the loading factor is positive: $\varepsilon \equiv \frac{\lambda}{2} - \frac{\gamma}{v} > 0$. In this case $p > q$ and

$$\mathbb{M}_\infty(x) = \frac{1}{\varepsilon}(1 + \gamma x).$$

Acknowledgements

The author acknowledges support from MICINN under contract No. MTM2009-09676.

References

[1] FELLER W. Integro-differential equations of purely discontinuous stochastic processes. *Trans. Amer. Math. soc.* 48 (1948), 488–515.

[2] R. N. BHATTACHARYA. *Stochastic Processes with applications*. John Wiley and Sons, New York, 1990

[3] DOOB, J. L. *Stochastic Processes.* John Wiley and Sons, New York, 1953

[4] COX DR. *Renewal Theory.* John Wiley and Sons, New York, 1965.

[5] KARLIN S, TAYLOR H. *A first course in stoch. processes*. Acad. press, N.Y. 1981.

[6] VERE-JONES D. Earthquake prediction, an statistician's view, *J. Phys. Earth 26*, (1978), 129–142.

[7] HELMSTETTER A, SORNETTE D. Diffusion of epicenters of earthquake aftershocks and continuous-time random walk models *Physical Review E 66*(2001)061104.

[8] RODRIGUEZ-ITURBE I, COX DR, ISHAM V.A point process model for rainfall: further developments 1988 *Proc. R. Soc. Lond., A 417* , 283–298.

[9] COWPERTWAIT P., COX DR, ONOF C. A point process model for rainfall. *Proc. R. Soc. Lond., A 463*, (2007), 2569–2587.

[10] V. S. OSKANIAN AND V. YU. TEREBIZH. Some characteristics of the flare activity of UV Cet Type stars. *Astrofizika 7,no. 1* (1971), 48–54.

[11] MERTON R.C. Option Pricing when underlying stock returns are discontinuous, *Journal of Financial Economics 3* (1971), 125–144

[12] CRAMER, H. Mathematical Th. of Risk. Skandia Jubilee Volume, Stockholm, 1930.

[13] LUNDBERG, F. Some supplementary researches on the collective risk theory. *Skandinavisk Aktuarietidskrift 15* (1932), 137–158.

[14] Sparre, ANDERSON E. On the collective theory of risk in the case of contagion between claims. *Transactions XVth International Congress of Actuaries 2*, (1957), 219–229.

[15] GERBER H.U. Martingales in Risk Theory, *Bulletin of the Swiss Association of Actuaries*. (1973), 205–216.

[16] GERBER H.U. AND SHIU, E.S.W.Actuarial Bridges to Dynamic Hedging and Option Pricing. *Insurance: Mathematics and Economics 18* (1996), 183–218

[17] DICKSON, D.C.M. AND C. HIPP. On the time of ruin for Erlang(2) risk processes. *Insurance: Mathematics and Economics, 29* (2001), 333–344.

[18] J. VILLARROEL AND M. MONTERO. Poisson driven stochastic process and nonlinear Schrodinger equation. *Stud. Appl. Math. 127(4)* (2011), 372–393.

[19] J. VILLARROEL AND M. MONTERO. On the effect of random inhomogeneities in Kerr media modelled. *J. Phys. B Atom. Molec. Opt. 43* (2010), 135404.

[20] S. LI AND J. GARRIDO. On ruin for the Erlang(n) risk process, *Insur. Math. Econ. 34*(2004), 391–408

Javier Villarroel
Facultad de Ciencias, Universidad de Salamanca,
Plaza Merced s/n, 37008 Salamanca, Spain
javier@usal.es

OTHER COMMUNICATIONS

The following contributions are the ones which were presented but not included in this book. Some will appear in other publications.

Modelling a bivariate counting process. An application to the occurrence of extreme heat events
Jesús Abaurrea, Jesús Asín and Ana C. Cebrián

Development of a daily to hourly rainfall disaggregation model
J. Abaurrea, J. Asín, A.C. Cebrián and N. Gavín

On the convergence rates of multivariate higher-order polynomial kernels
B. A. Afere and F. O. Oyegue

A heuristic for identifying unknown agents in a transaction network
David Alcaide, Joaquín Sicilia and Miguel Ángel González Sierra

Labour accessibility and residence attractiveness: a Bayesian analysis based on Spatial nteraction models
María Pilar Alonso, Asunción Beamonte, Pilar Gargallo and Manuel Salvador

Optimal Experimental Designs for Time Series Models
Mariano Amo-Salas and Jesús López-Fidalgo

Robustness and Density Power Divergence Measures
Ayanendranath Basu, Abhijit Mandal, Nirian Martin and Leandro Pardo

Asymptotic normality through factorial cumulants and partitions identities
Konstancja Bobecka, Paweł Hitczenko, Fernando López-Blázquez, Grzegorz Rempała and Jacek Wesołowski

Semiparametric estimation of a two-components mixture of regression models
Laurent Bordes, Ivan Kojadinovic and Pierre Vandekerkhove

Extending the Classical Koziol-Green model by using a copula function
Roel Braekers and Auguste Gaddah

Classification trees for the characterization of some aragonaise grapevine varieties
José Casanova, Beatriz Lacruz and Jesús M. Ortiz

Two perspectives of design and modeling when some independent variable is uncontrollable.
Víctor Casero-Alonso and Jesús López-Fidalgo

Free Completely Random Measures
Francesca Collet, Fabrizio Leisen and Antonio Lijoi

Likert and fuzzy scales: an empirical comparison through the MSE
Sara de la Rosa de Sáa, María Teresa López and María Asunción Lubiano

Statistical inference based on restricted sequential order statistics for weibull distribution with a power trend model
M. Doostparast and E. Velayati Moghaddam

Analysis of Demographic and Health Survey Data of Turkey with Bayesian Networks and Association Analysis
Derya Ersel and Suleyman Gunay

Optimum Designs for Enzyme Inhibition Models: Algorithmic Approach
Mercedes Fernández-Guerrero, Raúl Martín-Martín and Licesio J. Rodríguez-Aragón

An study of some frailty models
J.M. Fernández-Ponce, F. Palacios-Rodríguez and R. Rodríguez-Griñolo

Maximum likelihood and bayesian estimates and prediction for geometric distribution based on δ-records
R. Gouet, F.J. López, L.P. Maldonado and G. Sanz

Geometric Records
Raúl Gouet, F. Javier López and Gerardo Sanz

On the structure of near-record values
Raúl Gouet, F. Javier López and Gerardo Sanz

A least squares estimation method for high heteroscedastic linear models
Ali S. Hadi, Beatriz Lacruz and Ana Perez-Palomares

Departamento de
Metodos Estadísticos
Universidad Zaragoza

Instituto Universitario de Investigación
Biocomputación y Física
de Sistemas Complejos
Universidad Zaragoza

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 +302 pp., ISBN: 978-84-15031-53-6.

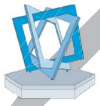
36. L. M. Esteban, B. Lacruz, F. J. López, P. M. Mateo, A. Pérez Palomares, G. Sanz y C. Paroissin (eds.): *The Pyrenees International Workshop on Statistics, Probability and Operations Research: SPO 2009*, 2010, 164 pp., ISBN: 978-84-15031-92-5.

37. J. Giacomoni, M. Madaune – Tort, C. Paroissin, G. Vallet, M. C. López de Silanes, M. Palacios, G. Sanz y J. J. Torrens (eds.): *Eleventh International Conference Zaragoza–Pau on Applied Mathematics and Statistics*, 2012, xxvi +208 pp., ISBN: 978-84-15538-15-8.

ISBN 978-84-15770-81-7

Prensas de la Universidad
Universidad Zaragoza

monografías
García de Galdeano



Instituto Universitario de Investigación
de Matemáticas
y Aplicaciones
Universidad Zaragoza