

ASMDA 2013

Book of Abstracts

15th Applied Stochastic Models and Data Analysis International Conference

With Demographics 2013 Workshop

Editor

Christos H. Skiadas



25 – 28 June 2013

Mataró (Barcelona), Spain

Imprint

Book of Abstracts of the 15th Applied Stochastic Models and Data Analysis International Conference (ASMDA2013), 25-28 June 2013, Mataró (Barcelona), Spain

Published by: ISAST: International Society for the Advancement of Science and Technology.

1st ed.

Editor: Christos H Skiadas

ISBN: 978-618-80698-2-4

© Copyright 2013 by ISAST: International Society for the Advancement of Science and Technology.

All rights reserved. No part of this publication may be reproduced, stored, retrieved or transmitted, in any form or by any means, without the written permission of the publisher, nor be otherwise circulated in any form of binding or cover.

Introduction

XVth Applied Stochastic Models and Data Analysis (ASMDA2013) International Conference

25 - 28 June 2013, Mataró (Barcelona), Spain

Since 1981, ASMDA aims to serve as the interface between Stochastic Modeling and Data Analysis and their real life applications particularly in Business, Finance and Insurance, Management, Production and Reliability, Biology and Medicine.

Our main objective is to welcome papers, both theoretical or practical, presenting new results having potential for solving real-life problems. Another important objective is to present new methods for solving these problems by analyzing the relevant data. Also, the use of recent advances in different fields will be promoted such as for example, new optimization and statistical methods, data warehouse, data mining and knowledge systems, computing-aided decision supports and neural computing.

The role played by ASMDA as interface between theory and practice means that the conferences are of great interest for both the academic and business world. The high standard of the meetings are guaranteed by strong international scientific committees.

The preceding international ASMDA Symposia were organized in Brussels (1981, 1983, 1985), Belgium, in Nancy (1988), France, in Granada (1991), Spain, in Chania (1993), Greece, in Dublin (1995), Ireland, in Anacapri (1997), Italy, in Lisboa (1999), Portugal, in Compiègne (2001), France, in Brest (2005), France, in Chania Crete (2007), Greece, in Vilnius (2009), Lithuania, and in Rome (2011), Italy .

In general, the proceeding volumes were published by international scientific publishers or local publishers.

The ASMDA 2013 Conference focuses on new trends in theory and applications of Applied Stochastic Models and Data Analysis.

A special Demographics2013 Workshop is organized under the umbrella ASMDA International. The aim is to gather people interested in improving demography and the related fields of analysis and research

including life and physical sciences as well as medical and technical information. We strongly support interdisciplinary studies and the improvement of the analytic tools and research methods

We thank all the contributors to the success of this conference, the ISAST Committee, the Secretary Mary Karadima, and especially the authors of this Book of Abstracts.

May, 2013

Raimondo Manca, Rome, Italy

Christos H. Skiadas, Chania, Crete, Greece

Vladimir Zaiats, Barcelona, Spain

ASMDA2013 Co-Chairs

ASMDA-2013 Committees

Chairman

Jacques Janssen, Honorary Professor of Université Libre de Bruxelles

Members

Alexander Andronov, Transport and Telecom. Institute, Riga, Latvia
 Vladimir Anisimov, Research Statistics Unit. GlaxoSmithKline
 Soren Asmussen, University of Aarhus, Denmark
 Helena Bacelar-Nicolau, University of Lisbon, Portugal
 Narayanaswamy Balakrishnan, McMaster University, Canada
 Elena Bacelar Nicolau, University of Lisbon, Portugal
 Paolo Baldi, University of Rome "Tor Vergata", Italy
 Vlad Stefan Barbu, University of Rouen, France
 Mireille Bardos, Banque de France
 Henry W. Block, Department of Statistics, University of Pittsburgh, USA
 Mark Brown, The City College, CUNY, USA
 Ekaterina Bulinskaya, Moscow State University, Russia
 Jorge Caiado, Centre Appl. Math, Econ, Techn. Univ. Lisbon, Portugal
 Enrico Canuto, Dipart. di Automatica e Inform, Politec. di Torino, Italy
 Manolis Christodoulou, Technical University of Crete, Chania, Greece
 Erhan Çinlar, Princeton University, USA
 Maria Mercè Claramunt, Barcelona University, Spain
 Marco Dall'Aglia, LUISS Rome, Italy
 Guglielmo D'Amico, University of Chieti and Pescara, Italy
 Pierre Devolder, Université Catholique de Louvaine, Belgium
 Giuseppe Di Biase, University of Chieti and Pescara, Italy
 Yiannis Dimotikalis, Technological Educational Institute of Crete, Greece
 Valerie Girardin, Université de Caen Basse Normandie, France
 Joseph Glaz, University of Connecticut, USA
 Gerard Govaert, Université de Technologie de Compiègne, France
 Alain Guenoche, University of Marseille, France
 Montserrat Guillen University of Barcelona, Spain
 Steven Haberman, Cass Business School, City University, London, UK
 Diem Ho, IBM Company
 Emilia Di Lorenzo, University of Naples, Italy
 Aglaia Kalamatianou, Panteion Univ. of Political Sciences, Athens, Greece
 Udo Kamps, Inst. für Stat. und Wirtschaftsmath., RWTH Aachen, Germany
 Alex Karagrigoriou, University of Cyprus, Cyprus

Włodzimierz Klonowski, Lab. Biosign. An. Fund., Polish Acad of Sci, Poland
 Tõnu Kollo, Institute of Mathematical Statistics, Tartu, Estonia
 Krzysztof Kołowrocki, Depart. of Math., Gdynia Maritime Univ., Poland
 Dimitrios G. Konstantinides, Dept. Stat. & Act. Sci., Univ. Aegean, Greece
 Volodymyr Koroliuk, University of Kiev, Ukraine
 Raman Kumar Agrawalla, Tata Consultancy Services, India
 Stéphane Lallich, University of Lyon, France
 Ludovic Lebart, CNRS and Telecom France
 Claude Lefevre, Université Libre de Bruxelles, Belgium
 Mei-Ling Ting Lee, University of Maryland, USA
 Philippe Lenca, Telecom Bretagne, France
 Nikolaos Limnios, Université de Technologie de Compiègne, France
 Bo H. Lindqvist, Norwegian Institute of Technology, Norway
 Brunero Liseo, University of Rome, Italy
 Fabio Maccheroni, Università Bocconi, Italy
 Claudio Macci, University of Rome "Tor Vergata", Italy
 Raimondo Manca, University of Rome "La Sapienza", Italy
 Domenico Marinucci, University of Rome "Tor Vergata", Italy
 Laszlo Markus, Eötvös Loránd University - Budapest, Hungary
 Sally McClean, University of Ulster
 Gilbert MacKenzie, University of Limerick, Ireland
 Terry Mills, Loddon Mallee Integrated Cancer Service, Australia
 Leda Minkova, Dept. of Prob., Oper. Res. and Stat. Univ. of Sofia, Bulgaria
 Ilya Molchanov, University of Berne, Switzerland
 Karl Mosler, University of Koeln, Germany
 Annamaria Olivieri, University of Parma, Italy
 Enzo Orsingher, University of Rome "La Sapienza", Italy
 Valentin Patilea, ENSAI, France
 Mauro Piccioni, University of Rome "La Sapienza", Italy
 Ermanno Pitacco, University of Trieste, Italy
 Flavio Pressacco, University of Udine, Italy
 Yosi Rinott, The Hebrew University of Jerusalem, Israel
 Leonidas Sakalauskas, Inst. of Math. and Informatics, Vilnius, Lithuania
 Werner Sandmann, Dept. of Math., Clausthal Univ. of Tech., Germany
 Gilbert Saporta, Conservatoire National des Arts et Métiers, Paris, France
 José M. Sarabia, Department of Economics, University of Cantabria, Spain
 Sergio Scarlatti, University of Rome "Tor Vergata", Italy
 Hanspeter Schmidli, University of Cologne, Germany
 Dmitrii Silvestrov, University of Stockholm, Sweden
 Christos H. Skiadas, Technical University of Crete, Greece
 Charilaos Skiadas, Hanover College, Indiana, USA
 Dimitrios Sotiropoulos, Techn. Univ. of Crete, Chania, Greece
 Fabio Spizzichino, University of Rome "La Sapienza", Italy

Gabriele Stabile, University of Rome "La Sapienza", Italy
Valeri Stefanov, The University of Western Australia
Anatoly Swishchuk, University of Calgary, Canada
Andrea Tancredi, University of Rome "La Sapienza", Italy
Mariano Valderrama, University of Granada, Spain
Panos Vasiliou, Aristotle University of Thessaloniki, Greece
Larry Wasserman, Carnegie Mellon University, USA
Wolfgang Wefelmeyer, Math. Institute, University of Cologne, Germany
Shelly Zacks, Binghamton University, State University of New York, USA
Vladimir Zaiats, Universitat de Vic, Spain

Organizing Committee

Chairmen

Raimondo Manca

University of Rome "La Sapienza" Italy

Christos H. Skiadas

Technical University of Crete, Greece

Vladimir Zaiats

Escola Politècnica Superior, Universitat de Vic, Spain

Plenary Speakers

Narayanaswamy Balakrishnan

Department of Mathematics and Statistics
McMaster University
Hamilton, Ontario, Canada

Leonid Gavrilov

Center on Aging,
NORC at the University of Chicago
USA

Joseph Glaz

Department of Statistics (Chair)
University of Connecticut
Storrs, CT, USA

Steven Haberman

Cass Business School (Dean)
City University London
106 Bunhill Row, London, UK

Yury A. Kutoyants

Laboratoire de Statistique et Processus
Université du Maine
Le Mans, France

Pere Puig

Department of Mathematics, Group of Mathematical Statistics.
Universitat Autònoma de Barcelona, Cerdanyola del Vallés (Barcelona)
Spain

Jean-Marie Robine

Head of the research team Biodemography of Longevity and Vitality,
INSERM U710, Montpellier, France

Marco Scarsini

LUISS Guido Carli
Libera Università Internazionale degli Studi Sociali Guido Carli di Roma
Italy and
Singapore University of Technology and Design
Singapore

Dmitrii Silvestrov

Department of Mathematics
Stockholm University
Sweden

Contents	Page
Preface	iii
Scientific Committee	v
Organizing Committee	vii
Plenary Speakers	viii
Abstracts	1
Author Index	231

BOOK OF ABSTRACTS
Applied Stochastic Models and Data Analysis
ASMDA 2013 & DEMOGRAPHICS 2013

Plenary and Keynote Talks

One-Shot Device Testing and Analyses

Narayanaswamy Balakrishnan

Department of Mathematics and Statistics, McMaster University,
Hamilton, Ontario, Canada

Many devices like air bags in car and electro-explosive devices are tested for their reliability under one-shot device testing. In such a situation, by considering an exponential distribution for the lifetimes, I will describe the EM-algorithm approach for the determination of the maximum likelihood estimates of the model parameters under a single stress model. I will also describe a Bayesian approach for this case. Next, the results will be developed for the case of multiple-stress factors. The results will then be developed in more general terms by considering the Weibull lifetime distribution. Finally, some simulation results will be presented along with a real-life example to illustrate all the inferential results developed.

Biodemography of mortality and longevity

Gavrilov L.A., Gavrilova N.S

Center on Aging, NORC at the University of Chicago, USA

There is growing interest in scientific explanations of aging and in the search for a general theory that can explain what aging is and why and how it happens. There is also a need for a general theoretical framework that would allow researchers to handle an enormous amount of diverse

observations related to aging phenomena. Theoretical analysis of systems failure in aging leads naturally to apply the already existing general theory of systems failure, which is also known as the reliability theory. This lecture reviews existing theoretical reliability models and approaches, which help to understand the mechanisms and age-dynamics of systems failure. Empirical observations on systems failure in aging are also reviewed (the Gompertz and Weibull mortality laws, the compensation law of mortality, and the late-life mortality levelling-off), and are theoretically explained through the observed decline in system's redundancy with age. It is shown that redundancy is a key notion for understanding aging, and the systemic nature of aging in particular. Living organisms seem to be formed with a high load of initial damage, and therefore their life span and aging patterns may be sensitive to early-life conditions that determine this initial damage load during early development. Aging is a complex phenomenon, and a holistic approach using reliability theory may help analyze, understand, and, perhaps, control it.

Additional reading:

Gavrilov LA, Gavrilova NS. Models of Systems Failure in Aging. In: P Michael Conn (Editor): Handbook of Models for Human Aging, Burlington, MA: Elsevier Academic Press, 2006. 45-68. ISBN 0123693918

Full text available at:

<http://longevity-science.org/Failure-Models-2006.pdf>

Scan Statistics and Applications: Recent Advances and Future Developments

Joseph Glaz

Department of Statistics, University of Connecticut, Storrs, CT 06269,
USA

In the last five years, the area of scan statistics has emerged as one of the active and eminent areas of research in the field of applied probability and statistics. It has attracted leading researchers and scientists in diverse areas of science and technology. In addition to challenging computational and theoretical problems, the area of scan statistics has numerous applications in important areas of science and technology, including: astronomy, bioinformatics, biosurveillance,

engineering sciences, epidemiology, genetics, health sciences, material science, pharmaceutical sciences, physics, reliability and quality control and telecommunication (Glaz, J., Pozdnyakov, V. and Wallenstein, S. Scan Statistics: Methods and Applications, Birkhauser, Boston, 2009) . During the years 2008-2012, Google Scholar lists about 2,000 references, while the entire data base contains about 3,500 references on scan statistics. This year we are celebration half a century of modern era research in the area of scan statistics, originating in Naus (Clustering of Random Points in the Line and Plane, Ph. D. Thesis, Harvard University, Boston, MA, 1963). In this talk I will introduce scan statistics for one, two and three dimensional, discrete and continuous data. Both fixed and variable window scan statistics will be discussed. Approximations and inequalities for distributions of scan statistics will be presented for prospective, retrospective and sequential experiments. Applications in pharmaceutical sciences and engineering sciences will be discussed. Computational algorithms and simulations used in implementing and evaluating the performance of scan statistics will be discussed as well. Future research and several challenging problems in the area of scan statistics will be presented.

Keywords: Approximations, Inequalities, Local Detection, Monitoring, Moving Sums, Spatial Statistics.

Modelling and Projecting Mortality Improvement Rates

Steven Haberman

Cass Business School, City University London

The fact that we are living longer in many developed countries has a significant financial effect on individuals, governments, social security systems, pension plans and insurance companies. In order to plan in advance for these changes, we require reliable models that enable the forecasting of future trends and the measurement of uncertainty. In trying to understand trends in mortality rates over time, commentators have tended to use the idea of a mortality improvement rate as a descriptive tool. In this presentation, we investigate the direct modelling of mortality improvement rates and then the use of such models for forecasting. There are 2 broad approaches to defining a mortality improvement rate, using a period or a cohort perspective, and we illustrate both. We compare the results obtained with the more conventional approach to mortality projection, which involves the modelling of the mortality rate.

On ADF Goodness of Fit Tests for Stochastic Processes

Yury A. Kutoyants

University of Maine, Le Mans, France

We consider the problem of the construction of asymptotically distribution free tests for three models of stochastic processes: ergodic diffusion process, diffusion process with “small noise” and inhomogeneous Poisson process. For each model the basic hypothesis is supposed to be composite parametric and we propose the GoF tests based on statistics which have limit distributions “free of the underlying models”. These tests are similar to the well-known Cramèr-von Mises test.

Keywords: Goodness of fit tests, diffusion processes, Poisson processes, asymptotically distribution free tests.

Beyond the Maximum Likelihood characterization of distributions

Pere Puig

Department de Matemàtiques, Universitat Autònoma de Barcelona,
Bellaterra, Spain

The starting point of the maximum likelihood characterization of distributions was the well known result of K.F. Gauss which establishes that the only location model (under mild conditions) such that the sample mean is the Maximum Likelihood Estimator (MLE) of the location parameter, is the normal distribution.

Since then, there are many characterizations of distributions based on the Gauss’ principle, assuming that the MLE of the population mean is the sample mean. This corresponds to the generally accepted idea that the sample mean is the “best” estimator of the population mean. However, MLE has the good property of being asymptotically efficient, and this property can be used to characterize distributions as well.

Following this idea, in this talk we characterize all the symmetric location models for which the Laplace estimator, a linear combination of the median and the sample mean, is an asymptotically efficient estimator of the location parameter. The resulting model can be understood as a symmetrized or double truncated normal distribution.

Another example based in this property is the Logistic distribution. It is well known that the Hodges–Lehmann estimator is asymptotically efficient for the location parameter of the logistic distribution. In this talk

we give a simple and direct proof that this property also characterizes the logistic between all the symmetric location distributions (under mild conditions).

Keywords: Gauss' principle; Asymptotically efficient estimator; Location and scale models; Symmetric location models.

Health expectancies in the European Union, 2005-2011

Jean-Marie Robine

Head of the research team Biodemography of Longevity and Vitality,
INSERM U710, Montpellier, France

A key priority for the European Union (EU) is to ensure its ageing population keeps as healthy as possible and it has set a target to increase the number of healthy life years (HLY) by 2 years from 2010 to 2020. The latest trends show:

- In the EU as a whole life expectancy at age 65 in 2011 was 18.0 years for men and 21.4 years for women, an increase of more than a year from 2005 (1.3 years for men and 1.2 years for women).
- Life expectancy in good perceived health (years lived where people perceive themselves in good or very good health) has also significantly increased since 2005, by 1.5 years for men and 1.6 years for women, more than the increases in life expectancy.
- On a less positive note healthy life years (years free of activity limitation) at age 65 have remained stable between 2005 and 2011, increasing by only 0.2 years, to 8.8 years for men, and decreasing by 0.2 years, to 8.6 years, for women. This means that the years lived with activity restriction have increased between 2005 and 2011.
- Life expectancy without chronic illness, significantly decreased between 2005 and 2010 but increased markedly between 2010 and 2011 so that the 2011 values (7.2 years for men and 8.0 years for women) were similar to those in 2005. Again this means that the number of years lived with chronic illness increased between 2005 and 2011.

These apparently contradictory results might be explained in part by better management of health problems and/or associated disability (activity limitations), so that the health problems and disability associated with increases in life expectancy at age 65 have not generated an increase lower quality of life. It is likely that people, better informed today about their actual health status, report more chronic health problems and at earlier stages.

On information distortion in online ratings

Marco Scarsini (joint work with **Omar Besbes**)

LUISS Guido Carli, Libera Università Internazionale degli Studi Sociali
Guido Carli di Roma, Italy and
Singapore University of Technology and Design, Singapore

Consumer reviews and ratings of products and services have become ubiquitous on the internet. This paper analyzes the implications of the sequential nature of reports on the underlying performance they represent. We consider a sequence of consumers arriving sequentially over time and reporting a grade for some service. Upon arrival to the system, the consumer develops a sincere rating and also observes the average of past ratings. She then provides a report based on this information. We analyze how different behavioral models of consumer reports affect the long-run average of the reports compared to the true rating of the service.

We examine a broad class of behavioral models, and mixtures thereof. We study the long-run behavior of statistics of sequentially declared ratings. Focusing on the long-run average of the reports, we show that the behavioral model might introduce biases; furthermore, while the worst-case biases can be arbitrarily large for some herding mechanisms, it is less than 25% for compensating mechanisms. In addition, the ordering of the long-run averages preserves stochastic order of the true rating. We also illustrate further fundamental differences between compensating and conforming behaviors; in particular there exist conforming behaviors that allow manipulations of the long-run average of sequentially declared ratings through a limited and finite number of fake review.

Optimal Stopping and Convergence for American Type Options

Dmitrii Silvestrov

Department of Mathematics, Stockholm University, Sweden

Lecture presents a survey of results from a new book with the above title. The book is devoted to stochastic approximation methods for rewards of American type options for multivariate modulated Markov price processes with discrete time. The classes of log-price processes under consideration include multivariate modulated Markov chains, modulated random walks, and various autoregressive models. General convergence results are presented, as well as their applications to space

skeleton approximations, tree approximations, and Monte Carlo based approximation algorithms for option rewards. Also, results related to studies of structure for optimal stopping domains are presented as well as results related to option reselling problem. Theoretical results are illustrated by results of experimental studies. Finally, connection with problems of optimal stopping and convergence for multivariate modulated Markov type price processes with continuous time is discussed.

Keywords: American option, Optimal stopping, Convergence of rewards, Space skeleton approximation, Tree approximation, Monte Carlo approximation, Reselling of options

Invited and Contributed Talks

Assessing the Estimation Performance of Two Jittered Particle Filters Applied to State Space Models with Non-stationary Dynamics

Lesly M. Acosta¹ and M. Pilar Muñoz²

¹ Department of Statistics and Operations Research, Polytechnic University of Catalonia (UPC), Barcelona, Spain

² Department of Statistics and Operations Research, Polytechnic University of Catalonia (UPC), Barcelona, Spain

Sequential Monte Carlo filters - known as Particle Filters - have become a widely used approach to handle the estimation of the states (filtering) and/or parameters (learning) of dynamic models expressed in state-space form. The popularly used analytical Kalman Filter proves its optimality when filtering linear and Gaussian state-space models. When departures from this ideal context occur, alternative approaches should be entertained. Since we deal with non-standard state-space models and aim to simultaneously estimate the states and parameters, we adopt the particle filtering methodology. Specifically, this work aims to assess the statistical and computational performance of our SIRJ (sampling importance resampling with jittering [1]) particle filter and the LW (Liu and West adapted sampling importance resampling [2]) particle filter variant, taken as a benchmark. To fulfill our goal, we proceed as follows: first, based on data generated from a state-space model with a non-stationary dynamics, the efficient performance of the filters studied is assessed via Monte Carlo experiments. Second, the estimation ability of the two studied particle filter variants is illustrated on a real data set coming from the finance area; a stochastic autoregressive volatility model of order one, SARV(1), is entertained. Finally, key issues within the particle filtering framework are addressed in both the simulations and the application. These issues are: sample degeneracy, sample impoverishment, the choice of a resampling strategy, the impact of the number of particles, the time series length and the potential impact of the shrinkage parameter chosen.

Keywords: Particle filter, estimation, nonlinear, non-stationary, state space model, resampling, Monte Carlo, sequential parameter estimation.

References

- 1.M.P. Muñoz, M.D. and Marquez and L.M. Acosta. Forecasting Volatility by means of Threshold Models. *Journal of Forecasting*, 26, 343-363, 2007.
- 2.J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In: *Sequential Monte Carlo Methods in Practice*. Editors: A. Doucet, N. de Freitas and D. Gordon. Pages 197-222. Springer Verlag, 2001.

Coupling Method for Asymptotic Analysis of Queues with Regenerative Input and Unreliable Server

L.G.Afanasyeva, E.E.Bashtova

Lomonosov Moscow State University, Russia

First we define a regenerative flow and describe its properties. Then a single-server queueing system with a regenerative input flow and an unreliable server is considered. Applying the coupling we establish the ergodicity condition and prove the limit theorem in heavy traffic situation (traffic coefficient tends to 1 from below). The asymptotic analysis of the super-heavy traffic situation is also realized.

Keywords: Regenerative Flow, Unreliable Server, Ergodicity, Limit Theorems.

Functional logistic regression with derivative component

Ahmedou Aziza^{1,2}, Marion Jean-Marie^{1,3}, Pumo Besnik^{1,2,4}

¹PRES UNAM, Universités de Nantes, ²UMR CNRS 6093, LAREMA,

³Institut de Mathématiques Appliquées, UCO, ⁴Agrocampus Ouest, France

The generalized linear models (GLM) has been object of numerous studies and furnished interesting methods to analyse data in many applications. The reader will find a complete presentation of those models and theirs application in the book by McCullagh and Nelder (1989). Ractliffe et al. (2002) studied the particular case of logistic regression where predictors are functions. We present in this paper an extension of this last model by introducing a derivative component as predictor. Indeed, in some applications the information contained in the derived curves improves the predictive quality of the model.

The functional logistic regression with derivative component (FLogitRD) is defined as follows. Let Y be some binary random variable and X a random variable taking values on a Sobolev space $W=W^{2,1}[0,1]$, the Hilbert space of functions X such that the derivative DX belongs to $L=L^2[0,1]$. The model is $Y = E(Y/X) + \varepsilon$ with $E(Y/X) := \text{logit}[\eta(X; \beta, \gamma)]$ and

$\eta(X; \beta, \gamma) = \langle \beta, X \rangle_W + \langle \gamma, DX \rangle_L$ where $\beta \in W$, $\gamma \in L$ are the unknown parameters of the model and $X \in W$.

We propose to estimate the functional unknown parameters β and γ based on a sample (Y_i, X_i) of size n . Clearly the classical methods can not be applied for such model and we propose to transform it into a multivariate GLM model by using B-spline basis. We prove some consistent properties of the estimators and obtain some convergence rates using results given in Stone (1986) and Cardot and Sarda (2005). We illustrate our results to “annual precipitation for Canadian weather stations” data (see Ramsey and Silverman, 1996) and compare numerically our results with the functional logistic regression without derivative component.

Keywords: functional data, glm, logistic regression, derivative component

References

- Cardot H., Sarda P., Estimation in Generalized Linear model for Functional Data via Penalized Likelihood, *J. of Mult. Data Analysis*, 2005
 McCullagh P., Nelder J., *Generalized Linear Models*, 2nd ed., Chapman: London, 1989.
 Ramsey J., Silverman B., *Functional Data Analysis*, Springer, 1996.
 Ratcliffe S.J, Leader L.R and Heller G.Z, Functional data analysis with application to periodically stimulated foetal heart rate data II: Functional logistic regression, *Statist. Med*, 2002
 Stone C.J., The dimensionality reduction principle for generalized additive models, *Ann. Stat.*, 1986, 14(2), 590-606.

An Unreliable Retrial Queue with Impatience and Preventive Maintenance

Amar Aissani¹, Samira Taleb¹, Djamel Hamadouche²

¹ University of Sciences and Technology USTHB, ² University UMMTO, Algeria

Retrial Queues are useful in the stochastic modeling of computer and telecommunication systems amongst others. In this paper we study some problems connected with retrial phenomena involving by the unreliability of the server line which can be subject to physical breakdowns or computer attacks. The model considered here takes into account the corrective and preventive maintenance. Such policies are useful for example in many practical situations as in production systems or for the maintenance of website subject to computer attacks. We obtain the joint probability distribution of the server state and the number of orbiting customers in the system. This distribution is obtained in terms of Laplace and z- transforms.

Keywords: Retrial queues, Unreliable Server, Piece-Wise Markov Process, Laplace and z transforms.

Prediction of Non-Methane Hydrocarbons in Kuwait Using Regression and Bayesian Kriged Kalman Model

F. A. AL-AWADHI

This article describes the hierarchical Bayesian approach for predicting average hourly concentrations of ambient non-methane hydrocarbons (NMHC) in Kuwait where records of six monitor stations located in different sites are observed at successive time points. Our objective is to predict the concentration level of NMHC in unmonitored areas. Here an attempt is made for the prediction of unmeasured concentration of NMHC at two additional locations in Kuwait.

We will implement a Kriged Kalman filter (KKF) hierarchical Bayesian approach assuming a Gaussian random field, a technique that allows the pooling of data from different sites in order to predict the exposure of the NMHC in different regions of Kuwait.

Keywords: space-time models, autoregressive processes, spatial interpolation, Kriging, Kalman Filter, hierarchical model, spatial prediction, non-methane hydrocarbons, air pollution.

Does Government Ownership and Disclosure affect Performance and Stability of the GCC Banking Sector?

Ritab Al-Khour

Qatar University, Qatar

This paper studies empirically the effect of government ownership and the level of disclosure on the performance and the stability of banks, controlling for regulations, concentration, bank and country specific characteristics. The sample used covers 59 banks in six countries which comprise the Gulf Cooperative Council (GCC), for the period from 2004 to 2010. To test our hypotheses, two methodologies are implemented: The first is the Generalized Least Square Random Effect (GLS RE) methodology; and the second is the Generalized Method of Moment (GMM) methodology, which controls for the endogeneity and the omitted variable problems that occur in this kind of studies. Contrary to expectations, we find that governmentally owned banks are more stable than their counterparts. However, there is very limited evidence on the negative relationship between government ownership and bank

performance. As documented in the literature, bank size is positively and significantly related to bank profitability and negatively related to risk taking. The level of disclosure negatively affect bank performance, indicating that investors may not view information disclosed by the bank as reliable or may not be able to interpret correctly the information disclosed. Therefore, this might have reduced the positive effect of disclosure on stability and bank valuation. In addition, the increase in the level of concentration of the banking sector lowers the bank stability. High concentration (measure of the level of competition) by banks is perceived negatively by investors and consequently more competitive environments are considered to permit greater stability in banking systems. Therefore, policy makers should consider the possibility of opening their markets for more foreign banks and other regional banks to improve the competition in the market. Finally, the impact of bank regulations on bank risk taking is negligible.

Spatial-Temporal Interpolation of Non-Methane Hydrocarbons Levels in Kuwait

Shafiqah Alawadhi, Fahima Alawadhi

Department of Statistics and Operations Research, Kuwait University,
Kuwait

In this work we describe an approach for predicting average hourly concentrations of ambient NMCH in Kuwait. We use hierarchical Bayesian approach. First we model the logarithmic field as a trend model plus Gaussian stochastic residual. That trend model depends on hourly meteorological predictors and is common to all sites along with the seasonality effect. The stochastic component consists of a 24 hour vector response that we model as a multivariate ARIMA temporal process with common spatial parameters. Removing the trend and ARIMA structure leaves "whitened" time series of vector series. With this approach there is little loss of spatial correlation in these residuals compared with that in just the detrended residuals. Then we develop a spatial predictive distribution for these residuals at unmonitored sites. By transforming the predicted residuals back to the original data scales we can impute Kuwait's hourly NMCH.

KEYWORDS: space-time models; autoregressive processes; spatial interpolation; monitoring networks; spatial correlation

**Limitations and assumptions of experience curves used as
an instrument to evaluate the competitiveness of
photovoltaic technologies in the energy market**

Heino Albering¹, Michal Greguš²

¹ Comenius University, Faculty of Management, Odbojárov 10, P.O. Box
95, 820 05 Bratislava 25, Slovak Republic

² Comenius University Faculty of Management, Odbojárov 10, P.O. Box
95, 820 05 Bratislava 25, Slovak Republic

Experience curves describe the reduction in unit production cost with an increase in cumulative production. This concept is used to examine past price decreases in photovoltaic [PV] and to assess the potential economic impact of different PV growth scenarios. This article makes research in the literature on PV experience curves and their application to energy technology.

Before 2000, there have been published a few studies on PV experience curves. But in recent years an increase of literature on experience curve theory for energy technology can be noted. In both, scientific literature and policy documents of experience curves are partially mixed with the learning curves. But there is an important distinction, as a rule, the learning curves analyze only one factor, while experience curves include factors such as the costs of research, management, marketing or capital; they consider the total costs and cumulative production analysis of the entire market.

This paper considers that care must be taken in the application of experience curves to evaluate the competitiveness and to forecast cost reductions in PV. In addition, we explain that special attention should be paid to the reference that are being used, because a lot of research was done on the development of the module prices, but only a little in the field of the experience curves of PV kWh [Kilowatt-hour] prices.

Keywords: Experience curves, energy technology, photovoltaic.

**Multiple testing for the detection of local dependence
between two point processes. Application to the study of
neuronal spike trains synchronization**

**Mélisande Albert, Yann Bouret, Magalie Fromont, Franck
Grammont, Patricia Reynaud-Bouret**

Laboratoire J. - A. Dieudonné Université de Nice Sophia Antipolis,
FRANCE

Considering an i.i.d. sample from the same distribution as a pair of point processes observed on a given time period, we address the question of detecting dependence time windows between the two underlying point processes.

As a first step, we propose several single tests of independence between the two point processes on one time window, whose test statistics are based on the so-called delayed coincidences count. Since in the present work, the point processes are not necessarily assumed to be Poisson processes, the choice of the corresponding critical values is a crucial and tricky question. We propose here to construct the critical values from permutation and bootstrap methods, proving that the resulting tests are exactly or at least asymptotically of prescribed level α . We also study the consistency of the tests against particular alternatives.

As a second step, we propose to combine these single tests applied on various sliding time windows, following the multiple testing approach of Benjamini and Hochberg (1995).

We finally evaluate the ability of our new multiple testing procedures to detect dependence time windows on simulated data, before applying them on real neuronal data collected from electrodes recordings of spike trains for pairs of neurons in a primate cortex during n experiments which are assumed to be independent and identical.

Keywords: Point processes, independence tests, permutation tests, bootstrap methods, multiple testing, neuronal spike trains, neuronal synchronization

Approximations for two-dimensional discrete scan statistics in some dependent models

Alexandru Amărioarei^{1,2}, Cristian Preda¹

¹Laboratoire de Mathématiques Paul Painlevé, UMR CNRS 8524, MODAL-INRIA Lille, EA 2694 Université des Sciences et Technologies de Lille, France, ²National Institute of R&D for Biological Sciences, Romania

We consider the two-dimensional discrete scan statistic generated by block factors from i.i.d. sequences. We present the approximation for the distribution of the scan statistics and error bounds. A simulation study illustrates our methodology.

Keywords: Scan statistics, m -dependent sequences, block-factor.

Social conditions and the Gompertz rate of ageing

Jon Anson

Dept. of Social Work, Ben Gurion University of the Negev, Israel and Visiting Fellow, Royal Flemish Academy of Belgium for Science and the Arts, Belgium

As living standards improve, mortality declines, and the age-pattern of human mortality changes. Yet, while much of human mortality decline can be related, directly and indirectly, to changes in the material

standards of living, differences in human mortality, and in the shape of the mortality curve, cannot be explained in terms of material standards of living alone. Different social conditions, levels of technology and patterns of social relations, between social groups and between men and women, will result in a different age structuring of mortality, even in populations with a similar average level of mortality. How, then, can we describe and explain these differences in the shape of the mortality curve?

We use a random sample of 109 tables drawn from the Human Mortality Database and consider what may be learnt from a parametric (Gompertz) representation of the underlying mortality curve above age 35. As the general level of mortality declines so does the level of mortality in young adulthood, and the proportion surviving to young adulthood increases. The slope (β) parameter represents the rate of ageing of the population and is thus positively correlated with life expectancy at birth (e_0) and survivorship to age 35 (l_{35}). However the specific rate of ageing will vary for any given level of e_0 or l_{35} , according to whether adult mortality occurs early or is delayed. Approximately half the variation in β is accounted for by the general level of mortality in the population. In the present paper we consider how the social conditions in which the population lives (standards of living, family structures, social inequality, etc.) affect the rate of population ageing for men and for women given a particular level of survivorship to young adulthood.

Keywords: Mortality, Life tables, Gompertz.

Methods and tools for nonlinear analysis of chaotic time series

Valery Antonov, Anatoly Kovalenko, Artem Zagaynov, Vu van Quang

Saint-Petersburg State Polytechnic University, Russia

The report reveals of the various methods of assessment, monitoring and forecasting of time series by calculating indices of correlation dimension, approximate entropy, Lyapunov exponent, Hurst exponent, etc. used in various software tools (eg application FracLab package MatLab, software RQA (aut. CJ Webber), application CorDimension (Copyright FRNMT AIRES), package Tisean (avt. R. Hegger et al.) package fractal 4.4 (avt. Sychev), etc.). It is shown the imperfection of numerical methods for the selection of the control parameters of the existing procedures calculated (the calculation of the scaling region, the required embedding dimension, delay parameters, the distances to calculate approximated entropy) and proposed their own methods of assessment. We have developed own software, methods that improve

on some of these computational parameters. This paper is a continuation of the subject, the authors repeatedly presented at a conference [1-2,4].

Keywords: correlation dimension, approximate entropy, Lyapunov exponent, Hurst exponent

[1] V. Antonov, J. Danilevich A. Kovalenko, A. Zagajnov Dynamic-entropy monitoring daily state human heart rates // Abstracts The XIII International Conference «Applied Stochastic Models and Data Analysis» (ASMDA-2009). Vilnius 2009. Vilnius «Technika» P. 8.

[2] V. Antonov, A. Kovalenko, A. Zagaynov, Vu van Quang The research of fractal characteristics electrocardiogram in real time // Book of Abstracts International Conference «Stochastic Modeling Techniques and Data Analysis» (SMTDA-2010). Chania Crete Grece 2010. MAICH Prefecture of Chania, P.11

[3] V. Antonov, A. Kovalenko, A. Zagaynov, Vu van Quang The research of fractal characteristics of the electrocardiogram in a real time mode // Journal of Mathematics and System Science. 2012. Vol. 2, №3, P. 191-195.

[4] V. Antonov, A. Fedulin, S. Nosyrev, A. Kovalenko Critical Condition in Human. New Approach To The Entropy Based Technology of Definition // Abstracts The XII International Conference «Applied Stochastic Models and Data Analysis» (ASMDA-2007). Chania Crete Grece 2010. MAICH Prefecture of Chania, P. 10.

[5] Antonov V., Fedulin A., Nosyrev S., Kovalenko A. Critical Condition in Human. The Entropy Based Technology of Definition // International Journal Communications in Dependability and Quality Management (CDQM – Journal) Vol. 10. #1, 2007

On some properties of geometric Poisson distribution

Masood Anwar

Department of Mathematics, COMSATS Institute of Information
Technology, Pakistan

The geometric Poisson (also called Pólya-Aeppli) distribution is a particular case of compound Poisson distribution. In this article we provide a recursive formula for computation of probabilities of geometric Poisson distribution (GPD) and prove that GPD is infinitely divisible, not log-concave, unimodal and also obtain its survival function. Finally its characterization is done via recursive relation of factorial moments.

Some Ranked Set Sampling Based Strategies for Statistical Process Control

Muhammad Rizwan Anwar

Control Charts are the best tools to monitor the stability and variation in every process. The shifts in the process may occur in location or dispersion parameter of the variable of interest.

There is a great development in the sampling strategies which result in to improved Statistical structures in general and efficient control charting structures in particular. The study was planned to investigate Ranked Set Sampling (RSS), Median Ranked Set Sampling (MRSS) relative to Simple Random Sampling (SRS) scheme for monitoring location and dispersion parameters using different run rule schemes particularly $\bar{\bar{x}}$ and $\bar{\bar{s}}$ control charts for location parameter; R and S control charts for dispersion parameters were considered. Power and Run Length (RL) properties were used as performance criterion to evaluate and compare the said control charting structures for location and dispersion monitoring using different sampling strategies and variety of run rule schemes. By considering bi-variate normally distributed process environment, it was found that ranked set sampling strategies have better ability to give out of control signals with the implementation of different run rule schemes both for location and spread parameters.

Approximation of the GI/M/1 Queue with Exponential Vacations

Sabrina Aoumer, Karim Abbas, Djamil Aïssani

Laboratory of Modeling and Optimization of Systems (LAMOS),
University of Bejaïa, Algeria

When modeling practical problems, the real system is often generally considered as a complex system which depends in complicated way on its parameters. Moreover, the parameters of the complex system are not often known exactly because they are obtained by statistical methods from empirical data. Therefore, the analysis of this type of complex systems doesn't allow us to obtain explicitly analytical formula. For this, one often replaces a complex system by another one which is close to it in some sense but more simpler in structure and/or component. However, in order to justify these approximations and estimate the resultant error, it is essentially of interest and important to precise the kind and the type of this perturbation and so the stability problem arises. Many variants of GI/M/1 system have been studied and applied extensively for performance evaluation purpose. In this talk we discuss the applicability of the strong stability approach to the GI/M/1 queueing

model with exhaustive service and multiple exponential vacations, and we obtain new results for the approximation of stationary characteristics of the GI/M/1 queue with vacations by those of the classical GI/M/1 queue.

Keywords: GI/M/1 queue, Vacation, Markov chain, Strong stability approach, Algorithm.

Effect of Anti-Malarial Medicines Use on the Quality of Health, a study in Kawempe division, Uganda, 2008

Innocent Asiimwe

Malaria kills more than a million people worldwide every year, with Sub-Saharan Africa (SSA) bearing about 90 percent of the global burden of malaria morbidity and mortality. In Uganda, the burden of disease due to malaria has remained unacceptably high, the main victims being children and pregnant women.

OBJECTIVES

The purpose of the study was to investigate anti-malarial medicine use in the community and its implications on the quality of health in Kawempe division.

METHODS

The study, cross-sectional in design, was carried out in Kawempe division. It was based on the hypothesis that poor quality health due to malaria in the division is a direct outcome of the irrational use of anti-malarial medicines. It employed quantitative data collection methods (structured questionnaires and semi-structured interviews) to obtain data on the variables affecting rational drug use and on those portraying the quality of health in the division. The categorical data was largely analyzed using the chi-square test by testing for differences of proportions for the independent variables in the study.

RESULTS

Only 31 (18%) of the respondents and 12 (40%) of the health workers knew that ACTs were the first line medicines for uncomplicated malaria. 86 (59.3%) respondents and 28 (93.3%) health workers would list more than three signs and/or symptoms of malaria. This led to a high number of respondents (73.3%, n=143) suffering from more than one episode of malaria in a year with the greatest proportion of those who were suffering the consequences of irrational use of anti-malarial medicines being those with lower levels of education. The variables that had significant associations with measures of quality of health (p-value less than 0.05) were: knowledge on anti-malarial drugs, capability to buy enough drugs, receipt and understanding of instructions, administering a complete dose and whether/not experienced side effects affected the way a respondent took his/her medication.

CONCLUSION AND RECOMMENDATIONS

The knowledge on ACTs was very low; ACTs were expensive and unavailable leading to irrational use of anti-malarial medicines since they, the ACTs, are the first line of treatment of uncomplicated malaria as stipulated by the Ministry of Health. There is need to increase public awareness of the change in malarial treatment policy and this should be accompanied by bringing treatment services closer to the people at as low rates as possible.

Conditional density estimation based on a clustering procedure

Stéphane Auray¹, Nicolas Klutchnikoff², Laurent Rouvière³

¹CREST-Ensai, EQUIPPE, ULCO and CIRPEE, ²CREST-Ensai, IRMA (UMR CNRS 7501), ³CREST-Ensai, IRMAR, UEB, France

Finite mixture models are widely used to account for population heterogeneities. In many fields such as biology, econometrics as well as social sciences, experiments are based on the analysis of a variable characterized by a different behavior depending on the group of individuals. A natural way to modeling heterogeneity is to use a mixture density model. This work deals with nonparametric estimation of conditional densities in mixture models. The proposed approach consists to perform a preliminary clustering algorithm to guess the mixture component of each observation. Conditional densities of the mixture model are then estimated using kernel density estimates applied separately to each cluster. The goal is to measure the effects of the clustering algorithm on the performances of the estimates of the conditional densities. To do that, we investigate the expected L1-error of the resulting estimates with regards to the performance of the clustering algorithm. In particular, we prove that these estimates achieve optimal rates over classical nonparametric density classes under mild assumptions on the clustering method used. Finally, we offer examples of clustering algorithms verifying the required assumptions.

Keywords: Nonparametric estimation, mixture models, clustering

Cluster Model Selection using Minimum Cost Spanning Trees

Renata Avros, Avi Soffer, Dvora Toledano-Kitai, Zeev Volkovich

Software Engineering Department, ORT Braude College, ISRAEL

Cluster model selection, often manifested as determination of the suitable number of clusters, is a crucial and ill-posed challenge in cluster analysis. Although many approaches have been proposed for solving this problem, none stands out as being the most effective for a general-

purpose model selection problem. Our approach presented here is based on the perception that samples drawn from the source population have to be well mixed in the clusters if the model is correctly chosen. Consequently, their clustered occurrences can be interpreted as independent realizations of the same random variable. A natural tool to measure the realizations' closeness is provided by simple probability metrics or two sample test multivariate statistics. In this paper we propose a method for cluster model selection, which is based on the minimum cost spanning tree approach. Our method is founded on a modified version of the Friedman and Rafsky two sample test which takes into account the edges' lengths. This Minimal Spanning Tree based test is intended to determining whether two sets of data originate from the same source. In our modified version of Friedman-Rafsky's algorithm, the statistic $J(x)$ uses not only the number of edges that connect points from different samples in a Minimal Spanning Tree that is constructed for each of the clusters, but also edge lengths in determining the suitability of a subset. Our approach for determining the number of clusters is summarized in a short Meta-algorithm. We then show how we adopted that meta-algorithm in the clustering process, in order to formulate our proposed approach to the cluster model selection. Finally, we present results of several numerical experiments which imply the ability of the suggested model to detect the true number of clusters.

Keywords: Clustering, Cluster Stability, Two sample test.

Empirical analysis of queues with abandonment: simple (but not too simple) models of complex systems

David Azriel, Paul D. Feigin, Avishai Mandelbaum

Faculty of Industrial Engineering and Management, Technion, Israel

Erlang-A is a simple birth-death stochastic process that has become popular for modeling call centers. ("A" stands for customers abandoning, or hanging up, due to impatience while waiting to be served). The Erlang-A model assumes, among other things, a fixed number of available agents, who serve customers from the queue whenever any customer awaits service. However, analysis of real data from call-centers, collected at the Technion SEELab, demonstrates that the latter assumptions are violated. Specifically, the number of agents that are available for service changes with system load and agents do idle while customers are waiting in queue. Moreover, such violations hurt the modeling and prediction power of Erlang-A.

This motivates a refinement of Erlang-A - a two-dimensional model with a large pool of agents, some of whom serve customers from the queue while others do not, and the tendency for serving customers increases with queue length.

The estimation of the new model parameters is challenging since the process is observed discretely in time, as opposed to its continuous evolution. We thus use an EM algorithm that computes the expected number of relevant quantities given our discrete-time data.

When comparing predictions of the two models against real data, we find that Erlang-A generally overestimates queue length and consequently the likelihood of abandonment, while our refined model predicts reality more closely and usefully.

Keywords: Call centers, Model selection, Queues with abandonment.

STUDY OF THE EURO CURRENCY FLUCTUATION BY USING DATA MINING

Maryam Baan, Elahe Saadati, Mahdi Nasiri

Iran University of Science and Technology, Tehran, Iran

Data mining or Knowledge Discovery in Databases (KDD) is a new field in information technology that emerged because of progress in creation and maintenance of large databases by combining statistical and artificial intelligence methods with database management.

Data mining is used to recognize hidden patterns and provide relevant information for decision making on complex problems where conventional methods are inefficient or too slow. Data mining can be used as a powerful tool to predict future trends and behaviors, and this prediction allows making proactive, knowledge-driven decisions in businesses. Since the automated prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools, it can answer the business questions which are traditionally time consuming to resolve. Based on this great advantage, it provides more interest for the government, industry and commerce.

In this paper we have used this tool to investigate the Euro currency fluctuation. This fluctuation depends on many different factors. For this investigation, we have used different algorithms and we have tried to extract Euro currency volatility by using the same criteria for all used algorithms.

The used dataset has 21,084 records and is collected from daily price fluctuations in the Euro currency in the period of 10/2006 to 04/2010.

Keywords: Euro Currency Fluctuation, Data Mining, Stock Market, Knowledge Discovery in Databases.

**Comparing time series segmentation techniques to detect,
locate and estimate change-points of heteroskedastic
processes**

Ana Laura Badagian, Regina Kaiser, Daniel Peña

Universidad Carlos III de Madrid, España

Time series segmentation has many applications in several disciplines as neurology, cardiology, speech, finance and others. Questions like: what are the main features of the brain activity when a epileptic patient suffers a seizure?; is the heart rate variability reduced after ischemic stroke?; what are the phonetic features to recognizing speech data?; is the conditional volatility of the financial assets constant?, can be answered performing segmentation analysis. The reason is that, many series in these fields do not behave as stationary but can be represented by approximately stationary intervals.

In this poster, we consider the problem of modeling a variance changing heteroskedastic time series, by segmenting the series into blocks which can be fitted by approximately stationary representation. The main objective involves describing, studying comparatively, and applying to real data, a number of segmentation or change-point detection techniques, including iterative cumulative sum of squares, Bayesian information criterion combined with binary segmentation, non-parametric Auto-SLEX, and parametric Auto-PARM. We compare the size in finite samples, the power, and the sensitiveness to the amount of change. We found that the size and the power of ICSS depends on whether or not the data are autocorrelated. The level of persistency also plays a essential role in this conclusion. BICBS, Auto-SLEX and Auto-PARM incorporate the data dynamic and have a better performance. The use of the procedures is illustrated with a finance time series dataset.

Keywords: Time series, change-points, segmentation

**A NONLINEAR MIXED EFFECTS MODEL TO EXPLAIN INTER-
INDIVIDUAL VARIABILITY IN PLANTS POPULATIONS**

Charlotte Baey, Samis Trevezas, Paul-Henry-Cournède

EPI DIGIPLANTE, INRIA Saclay Île-de-France, Laboratory of Applied
Mathematics and Systems, Ecole Centrale Paris, France

It is common knowledge that the genetic variability of plants, even of the same variety, can be very important and, if we add locally varying climatic effects, then the development of two neighboring similar plants could be highly different. This is one of the reasons why population-based methods for modeling plant growth are of great interest. A highly promising individual-based plant growth model is the GreenLab model (Cournède *et al.*[1]) which was recently shown to have a good predictive capacity among competing models (Baey *et al.*[2]). In this study, we

extend the GreenLab formulation to the population level. In order to model the deviations from some fixed but unknown important biophysical and genetic parameters we introduce into the GreenLab model appropriate random effects. Under some assumptions, the resulting model can be cast into the framework of nonlinear mixed effects models. A way to perform maximum likelihood estimation (MLE) for this type of models was presented in [3]. A stochastic variant of an EM-type algorithm (Expectation-Maximization) is generally needed to perform MLE for this type of incomplete data models and the interest is focused on the design of an efficient algorithm. In this direction, we propose and compare a suitable Monte-Carlo EM (MCEM) and a stochastic approximation EM (SAEM) algorithm for our model, where at each EM-iteration, MCMC is used to draw from the hidden states given the observed data (see also [4] for a MCEM algorithm for the individual plant growth model). It is noteworthy that with the proposed methodology, approximated confidence intervals can be easily obtained as a by-product of the algorithm. Several implementation issues, possible improvements and applications are discussed.

Keywords: plant growth model, mixed effects model, stochastic EM algorithms, MCMC methods, sugar-beet.

References

1. Cournède P.-H., Kang M. Z., Mathieu A., Barczi, J.F., Yan, H.P., Hu, B.G. and De Reffye, P. Structural Factorization of Plants to Compute their Functional and Architectural Growth. *Simulation* **82**(7), 427-438, 2006
2. Baey, C. Didier, A., Song, L., Lemaire, S., Maupas, F. and Cournède, P.-H. Evaluation of the Predictive Capacity of Five Plant Growth Models for Sugar Beet. *Proc. of 4th International Symposium on Plant Growth Modeling, Simulation, Visualization and Application*. Shanghai, 2012.
3. Kuhn, E. and Lavielle, M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, **49**, 1020–1038, 2005
4. Trevezas S., Malefaki S. and Cournède P. H. Simulation techniques for parameter estimation via a stochastic ECM algorithm with applications to plant growth modeling, (*submitted*).

Determination of the optimal strategy of a quarry in Algeria using the Three Phase Discrete-Event Simulation: A case study

Latifa BAGHDALI-OURBIH¹, Megdouda OURBIH-TARI², Abdelnasser DAHMANI²

¹Department of Infrastructures Base (DIB), Higher National School of Public Works (ENSTP), Kouba, Algiers 16000, Algeria

²Laboratory of Applied Mathematics University of Bejaia, Algeria

This work investigates how Turbo Pascal Three Phase Discrete- Event Simulation Routines and its libraries can be used in optimizing the

number of loaders and trucks in the aggregate production, at a minimum cost.

A model is developed using the activity cycle diagram and the performance measures of the system are evaluated. Inefficiencies of the studied problem are identified and improvement is proposed for the quarry managers by determining the optimal strategy.

Keywords: Modeling, Activity Cycle Diagram, Discrete-Event Simulation, Monte Carlo

Extreme value copulas and marginal effects

Zuhair Bahraoui, Catalina Bolancé, Ana M. Pérez- Marín

Riskcenter-University of Barcelona, Spain

In this paper we estimate the risk in an auto insurance company by the theory of copulas. This leads us to consider two fundamental issues. First we propose a goodness-of-fit test for extreme value copulas to selecting which the best copula reveals the dependence structure. The empirical results show that copula belongs to the family of extreme values. Second, we adjusted marginals based on the graphic evidence and we incorporate the Champernowne distribution function modified , wish gave a high capital requirement using the level required by Solvency II. Finally we treat to bounding the measure value at risk VaR for the aggregate loss by several methods in order to control the risk.

Keywords: Copula, VaR, Extreme Value

On the optimal designs for prediction of Ornstein-Uhlenbeck sheets

Sándor Baran¹, Kinga Sikolya¹, and Milan Stehlík²

¹ Faculty of Informatics, University of Debrecen, Kassai út 26, H-4028 Debrecen, Hungary

² Institut für Angewandte Statistik, Johannes Kepler University in Linz, Altenberger Strasse 69, A-4040 Linz a. D., Austria

Computer simulations are often used to replace physical experiments to explore the complex relationships between input and output variables. We study the optimal design problem for prediction of a stationary Ornstein-Uhlenbeck sheet on a monotonic set with respect to integrated mean square prediction error and entropy criterion. We show that there is a substantial difference between the shapes of optimal designs for Ornstein-Uhlenbeck processes (Baldi Antognini and Zagoraiou, 2010)

and sheets. In particular, we show that optimal prediction based on integrated mean square prediction error not necessary leads to space-filling designs. Moreover, we present some numerical experiments to illustrate selected cases of optimal designs for small number of sampling locations.

Keywords: Ornstein-Uhlenbeck sheet, Integrated mean square prediction error, Entropy, Fisher information.

Acknowledgments. This research has been supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund. This research was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system”. The project was subsidized by the European Union and co-financed by the European Social Fund.

References:

Baldi Antognini, A., Zagoraiou, M., 2010. Exact optimal designs for computer experiments via Kriging metamodeling. *J. Statist. Plann. Inference.* 140, 2607-2617.

Estimation of the entropy of semi-Markov chains

Vlad Stefan Barbu¹, Valérie Girardin²

¹ Laboratoire de Mathématiques Raphael Salem, UMR 6085, Université de Rouen, ² Laboratoire de Mathématiques N. Oresme, UMR6139, Campus II, Université de Caen Basse Normandie, France

A natural way to deal with the uncertainty of semi-Markov chains is to investigate their entropy, either the entropy of the stationary distribution or the entropy rate. When the process is observed, it becomes necessary to estimate this entropy.

The purpose of our talk is threefold: (i) to introduce semi-Markov chains up to deriving the associated entropy rate; (ii) to propose a plug-in estimator of the entropy rate of a semi-Markov chain; (iii) to investigate the asymptotic properties of the proposed estimator.

Keywords: homogeneous semi-Markov chains, statistical estimation, Shannon entropy, entropy rate.

References

V. Barbu, N. Limnios (2008) *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their use in Reliability and DNA Analysis*, Lecture Notes in Statistics, vol. 191, Springer, New York.

G. Ciuperca, V. Girardin (2007) Estimation of the Entropy Rate of a Countable Markov Chain, *Comm. Statist. - Theory and Methods*, 36, 2543–2557.

V. Girardin, N. Limnios (2004) Entropy rate and maximum entropy methods for countable semi-Markov chains, *Comm. Statist. Theory and Methods*, 33, 609-622.

A Tale of Two Investors: Estimating Risk Aversion, Optimism, and Overconfidence

Giovanni Barone-Adesi¹, Lorian Mancini², Hersh Shefrin³

¹Swiss Finance Institute and University of Lugano, ²Swiss Finance Institute and EPFL, ³Leavey School of Business Santa Clara University

We estimate investors' sentiment from option and stock prices by anchoring objective beliefs to a neoclassical pricing kernel. Our estimates of sentiment correlate well with other sentiment measures such as the Baker–Wurgler index, the Yale/Shiller crash confidence index, the Duke/CFO survey responses, and yet include additional information. Our analysis points out three significant issues related to overconfidence. First, the Baker–Wurgler index strongly reflects excessive optimism but not overconfidence.

Second, overconfidence drives the pricing kernel puzzle. Third, the dynamics of optimism and overconfidence generate a perceived negative risk-return relationship, while objectively the relationship is positive. Optimism and overconfidence about S&P 500 growth rates co-move together, inflating asset prices in good times and exacerbating market crashes in bad times.

Keywords: Sentiment, Risk Aversion, Pricing Kernel, Optimism, Overconfidence

JEL Codes: G02, G12

The Risk of Poverty Model in EU Population

Jitka Bartosova¹, Vladislav Bina¹, Marie Forbelska²

¹Department of Exact Methods, Faculty of Management, University of Economics in Prague, ²Department of Mathematics and Statistics, Faculty of Science, Masaryk University in Brno, Czech Republic

In the time of economic crisis the questions of poverty become of the great importance. In order to discover groups affected strongly by this problem it appears to be important to analyze factors influencing the situation at-risk-of-poverty. The contribution presents a generalized linear model (namely the multifactor logistic ANOVA model with mixed effects) of risk-of-poverty in EU countries in years 2004 to 2008. The

analysis is based on EU SILC survey data and focuses on the study of the impact of age group, sex, EU country and the year of survey on the risk-of-poverty.

Keywords: Multifactor Logistic ANOVA, Mixed Effects, Risk-of-poverty, EU SILC

Mean field optimal control games for jump processes

Rani Basna

School of Computer Science, Physics and Mathematics, Linnaeus University, Sweden

We investigate mean field games of the Markovian jump type with N players, who are weakly coupled via their empirical measure. We are assuming that the underlying space is finite. We show that the individual optimal strategy results from a consistent coupling of an optimal control problem with a forward nonlinear dynamics which leads to the well-known McKean-Vlasov dynamics in the limit as N goes to infinity. The limiting system represents a $1/N$ Nash Equilibrium for the approximating system of N players.

Keywords: Mean field, Markov jump process, Dynamic programming, Optimal control.

Asymptotic properties of trimmed sums

Alina Bazarova, Istvan Berkes, Lajos Horvath

Graz University of Technology, Austria

Trimming is a standard method to decrease the effect of large sample elements in statistical procedures to improve their efficiency. But Trimming is a standard method to decrease the effect of large sample elements in statistical procedures used, e.g., to construct robust estimators and tests. Trimming of i.i.d. sequences has been extensively studied from the 1960's and most basic problems of the theory have been solved, except a few isolated problems, e.g. the CLT under modulus trimming. In contrast, very little is known about trimming of dependent sequences, even though results here would be very useful e.g. in the statistics of heavy tailed processes. We formulate a few new results in this direction.

(a) We prove a functional CLT for trimmed AR(1) processes with stable errors, leading to a change point test for the unknown parameter of the process.

(b) We prove the CLT for trimmed \tilde{A} -mixing sequences, with applications in the theory of continued fractions.

Our method also gives insight into the central limit theory of modulus trimmed i.i.d. sums, showing that the difficulties in the classical theory can be removed by allowing random (but sample dependent) centering sequences in the CLT.

Keywords: stable distributions, trimming, change point, weak convergence, resampling

Comparison of fitting crop seedling emergence Models with some nonlinear model

Behnam Behtari

Department of Crop Ecology, Faculty of Agronomy, Tabriz Branch,
Islamic Azad University, Iran

Seedling emergence is one of the most important processes determining yield and the probability of crop failure. There has been accomplished numerous researches in recent years to understand and predict the emergence patterns of crop and weed species for different objectives. Nonlinear regression models have been developed to explain crop and weed emergence patterns as a function of time. In this study, we tested some seedling emergence models by field data of three crop species including green bean (*Phaseolus vulgaris* var. sunray), sunflower (*Helianthus annuus* L. var. Alistar) and maize (*Zea mays* L. var. Merit). Prediction of crop seedling emergence with the France and Thornley model and growth by the Logistic, Gompertz and Monomolecular models were also attempted. Emergence indices (SOE, MED, ERI, T0.5) showed that seedling emergence of maize was greater than green bean and sunflower. The values of the median emergence date (T0.5) predicted by Logistic model were in close agreement with the time required for 50% emergence calculated directly from interpolation of the raw emergence data. While shoot length (Lf) of crop seedling emergence fitted by Logistic and Monomolecular models were significant, it was not significant in Gompertz model statistically. Among the three models, the Gompertz and the Logistic models gave quite satisfactory results as the predicted values from the model and the observed values from the experiment were close (ME 0.9 in most of the cases and RMSE<8.0). The Monomolecular model was not suitable for prediction of studied crop seedling emergence due to its higher RMSE values (>12 in all cases). Results showed that the empirical models with an inflection point are recommendable because they predicted growth of crops seedling superiorly.

Keywords: Model Fitting, Modeling seedling emergence, Nonlinear regression

A generalization of some quantile-based measures commonly used in financial and insurance applications

Jaume Belles-Sampera, Montserrat Guillén, Miguel Santolino

Department of Econometrics, University of Barcelona, Spain

We propose a new family of risk measures -GlueVaR- which belongs to the class of distortion risk measures. This family is devised to reach a risk assessment lying between that provided by common quantile-based risk measures, namely value at risk and tail value at risk. A straightforward relationship between GlueVaR and these standard quantile-based risk measures is shown. It provides an easy procedure to obtain analytical closed-form expressions of GlueVaR for several distribution functions. Additionally, the concept of tail-subadditivity for distortion risk measures is introduced and investigated. A proof that a subfamily of GlueVaR risk measures satisfies tail-subadditivity is provided. An application to insurance claim data is presented.

Keywords: quantiles, subadditivity, tails, risk management.

A Characterization and Sufficient Conditions for the Total Time on Test Transform Order

Félix Belzunce, Carolina Martínez-Riquelme, José M. Ruiz

Universidad de Murcia, Facultad de Matemáticas, SPAIN

In this paper we provide a characterization of the total time on test (ttt) order, recently redefined by Hu, Wang and Zhuang (2012). This characterization provides a new interpretation of the ttt order and its dual, the dttt order. We also provide sufficient conditions for the ttt order when the stochastic order does not hold. Applications to the comparison of several parametric families of distributions are provided. We also suggest a potential application of the ttt and dttt orders for the comparison of risks.

Keywords: Total time on test transform, increasing concave order.

Comparing relative skewness of multivariate distributions

Félix Belzunce, Julio Mulero, José M. Ruiz and Alfonso Suárez-Llorens

Universidad de Murcia, Facultad de Matemáticas, SPAIN

In this talk we propose and study a comparison of relative skewness of multivariate distributions. These new criteria are based on the convexity of a certain transformation between the random vectors.

Keywords: Standrad construction, convex order.

Stationarity of Ornstein-Uhlenbeck processes with Stochastic speed of mean reversion

Fred Espen Benth¹, Asma Khedher²

¹Centre of Mathematics for Applications, Mathematics department, University of Oslo, Norway, ²Mathematical Finance, Technische Universität München, Germany

When modelling energy prices with the Ornstein-Uhlenbeck (OU) process, Barlow, Gusev, and Lai [1] studied the problem of estimating parameters based on historical data. They found out that there is a large uncertainty attached to the estimation of the speed of mean-reversion. In their study of Paris daily temperature data, Zapranis and Alexandridis [2] showed by means of wavelet analysis that the speed of mean-reversion is not constant but may vary considerably over time.

In this paper we generalised the OU process to allow for the speed of mean reversion to be stochastic. We analyze the stationarity properties of the mean and variance of this generalized OU process for simple specifications of the speed of mean-reversion.

We suppose that the speed of mean-reversion is a Brownian stationary process. Then, Despite the fact that the speed of mean reversion may attain negative values, we are able to show the stationarity of the mean and variance of the OU process when the average speed of mean-reversion is sufficiently larger than its variance. In our analysis, some aspects of Malliavin Calculus are applied.

Keywords: Ornstein-Uhlenbeck process, stationary process, Brownian stationary process, Malliavin calculus, chaos expansion.

References:

- [1] Barlow, M., Gusev, Y. and Lai, M.(2004). Calibration of multifactor models in electricity markets *International Journal of Theoretical and Applied Finance*, 7, (2), pp. 101-120
- [2] Zapranis, A., Alexandridis, A. (2009). Weather derivatives pricing: modelling the seasonal residual variance of an Ornstein-Uhlenbeck temperature process with neural networks.

On interrelations of generalized q -entropies and a generalized Fisher information, including a Cramér-Rao inequality

Jean-François Bercher

Laboratoire d'informatique Gaspard Monge, Université Paris-Est, France

In this communication, we describe some interrelations between generalized q -entropies and a generalized version of Fisher information.

In information theory, the de Bruijn identity links the Fisher information and the derivative of the entropy. We show that this identity can be extended to generalized versions of entropy and Fisher information. More precisely, a generalized Fisher information naturally pops up in the expression of the derivative of the Tsallis entropy. This generalized Fisher information also appears as a special case of a generalized Fisher information for estimation problems. Indeed, we derive here a new Cramér-Rao inequality for the estimation of a parameter, which involves a generalized form of Fisher information. This generalized Fisher information reduces to the standard Fisher information as a particular case. In the case of a translation parameter, the general Cramér-Rao inequality leads to an inequality for distributions which is saturated by generalized q -Gaussian distributions. These generalized q -Gaussians are important in several areas of physics and mathematics. They are known to maximize the q -entropies subject to a moment constraint. The Cramér-Rao inequality shows that the generalized q -Gaussians also minimize the generalized Fisher information among distributions with a fixed moment. Similarly, the generalized q -Gaussians also minimize the generalized Fisher information among distributions with a given q -entropy.

Keywords: Cramér-Rao inequality, generalized q -entropy, generalized Gaussians, de Bruijn identity.

Some New Stopping Rules for Multi-Stage Phase II Comparative Clinical Trials Involving Two Dependent Response Variables

Sotiris Bersimis¹, Takis Papaioannou¹, Athanasios Sachlas²

¹Department of Statistics & Insurance Science, University of Piraeus,

²Department of Nursing, University of Peloponnese, Greece

The aim of Phase II clinical trials is to determine whether an experimental treatment is sufficiently promising to justify further testing. If the experimental treatment gives positive evidence, the researchers proceed to Phase III trials. On the contrary, if more than expected adverse events appear, the researchers may interrupt the clinical trial earlier. Thus, appropriate stopping rules are needed for interrupting the clinical trial earlier than programmed either to proceed to a Phase III clinical trial or to abandon the experimental treatment. In this paper, we propose some new stopping rules that may be used for the early termination of clinical trials involving two dependent response variables.

These rules focus on stopping the clinical trial when we observe high rates of favorable or unfavorable outcomes with respect to a placebo or a competitive treatment, early enough. These new rules are defined on a multivariate sequence of outcomes originating from a multi-stage trial where at each stage two patients are involved (one receives the experimental treatment while the other receives the reference one) and after treatment they are tested with respect to two dependent characteristics. For the statistical justification of these rules, the distribution of the number of stages until certain types of events appear in the sequence is studied using the Markov Chain embedding technique. The numerical illustration showed a very good performance for these new rules.

Keywords: Bivariate Sequences of Multi-State Trials, Markov Chain Embeddable Random Variables, Multi-Stage Designs, Phase II Clinical Trials, Stopping Rules, Waiting Time.

A Meta-Heuristic Approach for the Optimal Deployment of Aerial Firefighting Fleets based on Predictive Fire Weather Risk Estimations

M. N. Bilbao¹, D. Gallo-Marazuela², S. Salcedo-Sanz², J. Del Ser³, C. Casanova-Mateo⁴

¹University of the Basque Country UPV/EHU, ²Universidad de Alcalá,
³TECNALIA RESEARCH & INNOVATION, ⁴Faculty of Sciences,
University of Valladolid, Spain

This paper sketches a novel meta-heuristic evolutionary approach for the optimal deployment of aerial firefighting aircrafts based on predictive fire risk estimations over a certain geographical areas. In particular, the problem can be formulated as how to properly allocate firefighting resources to capacity-constrained aerodromes in such a way that the utility of the deployed resources with respect to fire forest risk predictions is maximized. On the purpose of efficiently tackling this optimization problem, the proposed meta-heuristic solver inspires from the well-known Harmony Search algorithm, whose operators mimic the way musicians improvise new melodies in an aesthetically optimum fashion. Preliminary simulation results are presented and discussed, showing the effectiveness of the designed algorithm when handling problems of increasing dimensionality. This noted promising performance calls for new research lines towards considering further constraints and including new aspects for a more realistic problem formulation.

Keywords: Firefighting aircrafts, FWI, optimal deployment, Harmony Search.

Fractional Brownian Motion and its Application to Financial Time Series

Mária Bohdalová¹ and Michal Greguš¹

¹ Department of Information Systems, Comenius University in Bratislava,
Faculty of Management, Bratislava, Slovakia

The potential presence of stochastic long memory process in economic and financial time series has been an important subject of both theoretical and empirical research. The long-memory or long time dependence property describes the high-order correlation structure of a series. The lack of memory is checked by using R/S statistic and Lo's modified R/S statistic and adequate fractional Brownian motion (fBm) are simulated. The main contribution of this paper is fractal analysis of the selected financial time series.

Keywords: correlation, fractional Brownian motion, long memory process, simulation.

Dynamics of the many particle Jaynes-Cummings Model

Bogolubov N.N.(Jr.)¹, Rasulova M.Yu.^{2,3}, Tishaboev I.A.²

¹V.A.Steklov Institute of Mathematics of the Russian Academy of Sciences, Russia, ²Institute of Nuclear Physics Academy of Sciences Republic of Uzbekistan, Uzbekistan, ³MIMOS BHD, Technology Park Malaysia, Malaysia

We consider the dynamics of a system consisting of N two-level atoms interacting with a multi-mode cavity field, as an example of the generalized Jaynes-Cummings model [1]. Based on formulation of the collective atom variables the Jaynes-Cummings model is generalized to a system of N two-level atoms. For the given system, the generalized kinetic equation [2] is obtained and conditions are given under which its solution is reduced to solution of the linear equation, and of the nonlinear Schrödinger equation, respectively.

Keywords: Jaynes-Cummings model, nonlinear Schrödinger equation

References

1. Jaynes E.T., Cummings F.W., Proc. IEEE, 1963, **51**, 89.
2. Bogolubov N.N., Bogolubov N.N.Jr., Physics of Elementary Particles and Atomic Nuclei, 1980, **11**, 245, (in Russian); Bogolubov N.N., Bogolubov N.N.(Jr.), Aspects of Polaron Theory, Fizmatlit, Moscow, 2004, (in Russian).

Informative selection of a sample: asymptotic issues

Daniel Bonn  ry¹, F. Jay Breidt², Fran  ois Coquet¹

¹Ensai and IRMAR, France

²Department of Statistics, Colorado State University, Fort Collins, Colorado

In survey sampling, one cannot view the observations as independent ones for a number of reasons. First of all, one samples from a finite population; but also, the selection mechanism, i.e. the way one samples from this population, may induce specific dependence. As a consequence, the assumptions of most usual asymptotic tools of probability theory do not apply. There is however a need to properly address asymptotic issues when the size of the population one samples from is large, which means often.

In this talk, we shall consider informative selection of a sample from a finite population. Responses are realized as independent and identically distributed (i.i.d.) random variables with a probability density function (p.d.f.) f , referred to as the superpopulation model. The selection is informative in the sense that the sample responses, given that they were selected, are not i.i.d. f . In general, the informative selection mechanism may induce dependence among the selected observations. The impact of such dependence on the empirical cumulative distribution function (c.d.f.) is studied. An asymptotic framework and weak conditions on the informative selection mechanism are developed under which the (unweighted) empirical c.d.f. converges uniformly, in L_2 and almost surely, to a weighted version of the superpopulation c.d.f. This yields an analogue of the Glivenko–Cantelli theorem. A series of examples, motivated by real problems in surveys and other observational studies, shows that the conditions are verifiable for specified designs.

In the same context, we shall study asymptotic properties of the pseudo-maximum likelihood estimator in a parametric setting with a nuisance parameter.

Keywords: Survey sampling, informative selection, Glivenko-Cantelli Theorem, pseudo-likelihood.

Breast Cancer Survival at Braga's Hospital – Portugal

Ana Borges¹, Ines Sousa², Luis Castro³

¹Center for Innovation and Research in Business and Information Systems, Polytechnic Institute of Porto, ²Department of Mathematics and Applications of Minho's University, ³Senology Unit of Braga's Hospital, Portugal

The present work intends to describe the survival rate of breast cancer patients of the Senology Unit of Braga's Hospital, located in the north of

Portugal. Data consists in 451 female patients, diagnosed with a malignant tumour in the corresponding period of 1993 until 2012. We performed a survival analysis in order to describe the survival rate of these patients, as a function of possible risk factors, making use of Flexible Parametric Survival Models to estimate hazard ratios. Recurrence, type of surgical treatment, neoadjuvant treatment and triple negative breast cancer have a significant statistical effect on the global survival rate of these patients.

Results were also compared to those obtained when adjusting to the well known Cox proportional hazard model, and were quite similar for both models. There's an improvement of survival rate at 1, 3 and 5 years when compared to results from earlier studies.

Keywords: Breast Cancer, Survival, Flexible Parametric Survival model.

Graphical models for prognostics integration of complex discrete event systems

Mohammed-Farouk Bouaziz, Eric Zamaï

Laboratory of Grenoble for Sciences of Conception, Optimization and Production, France

These days, improved equipment availability is a critical factor in the competitiveness of manufacturers. The production facilities have very significant investment costs and they are operated at the limit of their capabilities. To improve equipments efficiency and maintenance activities, modeling methods must be able to define useable indicators in complex context.

The Equipment Health Factor (EHF) is a new indicator that can be used to predict the production equipment drifts. EHF is based on the analysis of historical operational data coming from: equipment parameters, maintenance events and products quality. Its calculation results using probabilistic graphical models (Bayesian networks BN) can be grouped into three classes of probability: good (G), medium (O) and bad (R) as shown in Figure 1.

The EHF integration consists in the use of the probabilistic results for maintenance (maintenance tasks management) and/or metrology (control steps management). For each phase an extension of the probabilistic model is necessary. In this paper we propose to use Petri nets method (PN) to simulate and evaluate several planning strategies in manufacturing workshops (e.g. semiconductor Fabs) based on equipment capacity changes following the EHF estimation. PN is represented by a static part (states, transitions and arcs) and a dynamic part (tokens) and it describe activities, resources and system states as shown in Figure 2. Also, the extended obtained model is a tool for

decision making. These efficient solutions allow the improvement of equipment availability, resource management, cycle time, costs, etc.

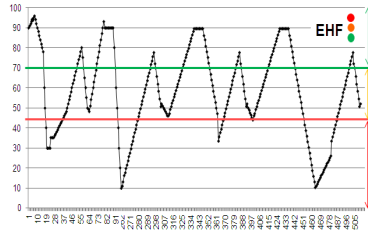


Fig. 1. EHF estimation

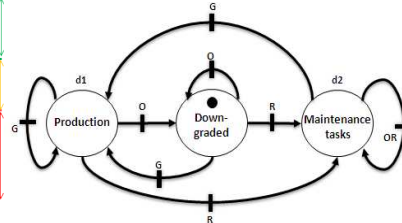


Fig. 2. Equipment monitoring

Keywords: Probabilistic graphical models, Petri nets, Prognostics and health management, Decision support, Resources management, Complex systems.

Convergence properties of mixture of factor analyzers with unequal number of factors

Charles Bouveyron, Julien Jacques

Université Paris 1 Panthéon-Sorbonne, France

The analysis and the clustering of high-dimensional data in nowadays a challenging and recurrent problem in many scientific fields. A popular and efficient way to model and cluster such data is to use factor analyzer models. Among the methods based on the factor analyzer model, we can cite mixture of PPCA (Tipping & Bishop), MFA (McLachlan et al.), PGMM (McNicholas & Murphy) and HDDC (Bouveyron et al.). Those models fit and classify the high-dimensional data in low-dimensional subspaces with a d factors. The inference of those models is usually done with the EM algorithm and the convergence properties is therefore guaranteed if the number of factors is fixed. Nevertheless, the assumption that the different subspaces have the same number of factors turns out to be too restrictive in some situations. Indeed, it seems natural to allow the subspaces to have different numbers of factors for fitting as well as possible the data of the different groups. However, the estimation of the factor numbers becomes a combinatorial problem if we do not allow them to vary along the iteration of the EM algorithm. Unfortunately, this relaxation implies the loss of the convergence guarantees given by the EM algorithm theory. In this work, we study the convergence properties of the mixture of factor analyzers when the number of factors is not fixed and can change over the iteration of the inference algorithm. In particular, we exhibit an evolution scheme for the

number of factors which guarantees the convergence toward a local maximum of the likelihood.

Keywords: model-based clustering, high-dimensional data, factor model, convergence properties

Calculating the Convexity Ratio and Applications

James R. Bozeman, Eva Glidden, Erin Milne, Timothy Nicholson

Lyndon State College, USA

The convexity ratio $CR(P)$ of a simple polygonal region P in the plane is equal to the (area of the endogon of P)/(area of the exogon of P). Here the endogon is a largest convex set contained in P . Such a set exists and it is polygonal, but it may not be unique (see [2] and [4] for the details). The exogon is the well-known convex hull of P , which is also polygonal. In this paper we present computer programs that calculate the endogon and its area given the number of vertices of P (see [3] and [4] for algorithms). We then present a program that calculates $CR(P)$. (Note that programs already exist for finding the convex hull of a polygon and its area.) If $CR(P) \geq .5$ we say that P is nearly convex. We apply this result to the shape of legislative districts, which are always polygonal, in the United States. A district is nicely shaped if $CR(P) \geq .5$, otherwise it is poorly shaped. A poorly shaped district may indicate partisan gerrymandering in any redistricting plan, which many states are trying to avoid (see [1] for the details). See [5] for a statistical analysis of an attempt to reduce the politics in redistricting by another method.

Keywords: Convexity ratio, nearly convex set, redistricting.

- [1] J.R. Bozeman, M. Pilling, The Convexity Ratio and Applications, *Scientiae Mathematicae Japonicae* (2013), (to appear).
- [2] J.R. Bozeman, L. Pyrik, J. Theoret, Nearly Convex Sets and the Shape of Legislative Districts, *Int. Jour. Pure and Appl. Math*, 49 (2008), 341-347.
- [3] J.S. Chang, C.K. Yap, A Polynomial Solution for the Potato-peeling Problem, *Disc. Comp. Geom.* 1 (1986), 155-182.
- [4] J.E. Goodman, On the Largest Convex Polygon Contained in a Non-Convex n -gon, Or How to Peel a Potato, *Geom. Ded.* 11 (1981), 99-106.
- [5] D.W. Peterson, Putting Chance to Work: Reducing the Politics in Political Redistricting, *Chance* 21(1) (2008), 22-26.

Modeling natural gas response to temperature**Marek Brabec, Ondrej Konar, Emil Pelikan, Ivan Kasanicky, Marek Maly**Department of Nonlinear Modeling, Institute of Computer Science,
Academy of Sciences of the Czech Republic, Czech Republic

In this paper, we will present a semiparametric regression model describing natural gas consumption for household customers and its dependence on ambient air temperature (as on its main driver). We will focus on daily average consumptions and daily mean temperatures. The model will be formulated on the level of an individual consumer and we will be interested both in the mean response and in the inter-individual variability of the response. Joint mean and variability modeling is useful for various balancing computations and estimates of their uncertainties. For obvious physical reasons, the consumption response is nonlinear, smoothly decreasing in temperature and having both lower and upper asymptotes. One possibility is to characterize its shape via a parametric model chosen as 1-cdf of an appropriately chosen distribution in the location-scale parametrization. Alternatively, we will proceed nonparametrically, using a B-spline basis here. There, we will explore dynamic character of the response, formulating a generalization of the now-classical Almon's distributed lag model frequently used in Econometrics and related fields.

Other, very important and complex feature of the consumption-to-temperature relationship is its time-varying character. At the simplest level, the effect of temperature can be seen as a (nonlinear) filter, but it is time-varying. In fact, response of the consumption to the same temperature (or even to its short time history) differs very much seasonally. We will address this problem in a careful state-space model formulation, where the temperature dependence structure will be allowed to change smoothly over position of a particular day within a year.

After describing the model structure and method for estimating its structural parameters (which will be essentially MLE), we will illustrate its performance on real large-scale data obtained from the Czech Standardized Load Profiles project. Namely, we will work there with a sample of several hundreds of household customers which have been followed for several years with gas meters recording in very fine time resolution. These essentially continuous data allow us to compare our previous approach to that based on functional data analysis methods (like e.g. the functional principal components analysis).

In addition to demonstrating the model performance, we will also illustrate one of the practical uses of a (simplified) model version for detection of customers with unusual temperature-corrected consumption trajectory which is a part of a system being developed at our department for RWE.

Demographics: quantity and quality too?

Grigori I. Brekhman

Israel, Russia

The analysts of the International Monetary Fund (IMF) in their presentation to the UN reported with concern: in the last 60 years the population of Earth is increased not only due to the birth rate, but also due to the increase in life expectancy. They predict the collapse of the financial system, because of increased spending on pensions, treatment and care for the elderly, if this trend will continue to 2050. The analysts, focusing their attention on the costs for retirees, did not discuss the huge costs for a born now and those who will be born in future. The birth rate is different in various countries; however not number of born children is a problem, but their quality - mental and physical. Why?

1. The statistic analysis of the morbidity in different countries shows that the health of pregnant women, usually poor. Diseases of the circulatory, endocrine, urinary, immune and other systems has found at many women. This means that children are born with a predisposition to diseases of like systems and organs (Shekhtman et al., 1982). These diseases are found in childhood, they have a chronic tendency and accompany the individual in lifetime, periodically become acute. They require ongoing management, and they are accompanied by reduce the mood and ability to work of individual.

2. Poor health of parents, their emotional distress, poor quality eggs - the cause of miscarriage and prematurity with birth of the children with the low weight, including children with extremely low birth weight. This requires lengthy resuscitation without the firm belief that the newborn babies save the good mental and physical health. The surviving children have often the delay of mental development, inadequate functioning of organs and systems, which are considered as a chronic diseases. They have the decreased ability to work, some of them need for a constant and long-term care. Simultaneously they have the trauma by everything that surrounded the child in intensive care. And this persists for lifetime on the unconscious level, defining their thinking and behavior in future.

3. Socio-economic conditions (hunger, poverty, war) promote the birth of children, who biologically different from those born under favorable circumstances. Severe conditions of development before birth can predispose to diseases of the cardiovascular system (coronary heart disease, hypertension) that occur at a younger age in the event of such situations (Barker et al. 2006).

4. It is appropriate to say that chronic diseases usually are not isolated. Multimorbidity - heavy burden for patients, the system of health care and insurance companies. Multimorbidity acquire the character of a psychosomatic process and involve in the diagnostics and treatment of 3-4 and more specialists. Even complex treatment of patients with

multimorbidity in the conditions of specialization modern medicine does not guarantee success, because of the lack of knowledge by physicians of the psychosomatic processes, their diagnostics and conservative treatment, as well as difficulties for removing the psychosocial factors that have a pathogenic effect on humans.

5. The unwanted pregnancy is an important factor in the violation of quality of the birth people. "The unwanted children" have a line of psychological peculiarities that bring sufferings both to them and surrounding people. They constantly experience the psychological discomfort, inferiority complex, prompting them to smoking, alcohol, drugs. As a rule, they have certain psychosomatic diseases or syndromes. Among people who have had the attempt to physically abortion them, or who have survived after having lived in an unwelcoming and ambivalent prenatal environment in which the possibility of their being aborted had been consciously or unconsciously of by their parents or others rejection after birth as confirmation of their undesired are a high frequency of aggression and auto-aggression (J.Sonne 2005). The latest example is the tragedy at the school in Connecticut (USA) in December 2012. Help: By the data of the WHO experts in 2003 in the world was executed more than 42 million of abortions. Frequency of abortions on the world on the average are 31 abortions on 100 born children, the highest frequency in East Europe - 105 abortions on 100 born children.

6. The birth trauma is discussed from 1924 (O.Rank). Recent studies have shown such chain of events: under the influence of external information most women have fear of labor and the fear of childbirth hospital. The emotional stress of mother and child is able to disrupt the natural birth. Of course, obstetricians have a large arsenal of tools to speed up and correction of parturition (pharmacological agents, various manipulations). Often such speedy way is a cesarean section (at the woman's request or at the obstetrician's suggestion), its frequency in different countries varies from 25 to 70%. Modern philosophy and technology of delivery, in fact, reflect the status of society with its accelerated rhythm of life, a pragmatic approach, psychological illiteracy and "emotional deafness." Today all of that surrounds the birth, even obstetricians called "obstetric aggression" (Radzinsky, 2010). All this causes trauma at both the mother and child (Grof 1975-2005).

7. The researches in the field of prenatal and perinatal psychology have shown that prenatal trauma at the child becomes a source of a variety of mental disorders: fear, phobia, psychomotor retardation, autism, syndrome ADHD, depression, bipolar disorders, schizophrenia and other. Prenatal trauma of an unborn child is a result of emotional stress, experiencing along with his mother, because of her psychological features and her way to respond to stressful situations. The latters often occurs under the influence of the nearest (nuclear family, extended family) and remote (school, work, religion, state) social surrounding. The

violence - domestic, street, clan, terror, wars - plays a special role in origin of prenatal trauma. It is manifested in depression, or, conversely, in very aggressive thinking and behavior (L. deMause 1998, L. Janus 2001).

This is a partial list of the factors that affect human before and after birth, remaining in his memory and promoting the emergence of new mutations. Much from the saved experience is passed to the following generation. The "vicious circle" already generated. How it to break off? What possible ways to avoid, to eliminate or at least reduce exposure destructive factors?

At the first stage it is necessary: a) wide informing of the population on mental life of the child before his birth, his emotional perception and actively functioning memory, b) about influence of the information received up to birth on a thinking and behaviour of the individual in the subsequent life, c) about the prenatal and perinatal sources of mental and somatic disorders at individuals, and d) an opportunity of their preventive maintenance.

Geo-Sectorial Interdependence Models for the Study of Credit Risk Evolution

Marina Brogi¹, Guglielmo D'Amico², Giuseppe Di Biase², Raimondo Manca¹, Giovanni Salvi¹

¹ Università di Roma "La Sapienza" Roma Italy

² Università di Chieti-Pescara Chieti Italy

The globalization of the world has been one of the engine for the development of the world economy. In the same time it is well known that a big economic crisis of a key country can have relevant consequences on the economies of other countries in the world.

The regional Input-Output models permits, having data, to study the interconnections among the economies of different countries. It is clear that the most the regions are connected the most their economic situations will be influenced. The aim of this paper is to investigate how the interdependence among different regions and, consequently, sectors are influenced by the rating evolutions.

At sectorial level, it is easy to think that if two sectors, in the same and/or in different countries, are strictly connected the rating evolution of one sector should influence the rating evolution of the other. We think that this sector influence is not symmetric in the sense that can happen for example that the default of one sector can give great problems to another sector but that the inverse relation could not have the same strong influence. In the same time we think that at regional level this fact is not true, in the sense that a crisis in one region can give problems to

another country and vice-versa. The last five year occurrences prove our assertion.

The rating migration models give the possibility to study the time evolution of ratings and, consequently, to obtain the probability of default given a starting rating at a given time. The regional Input-Output models give the interconnections among the sectors of different geographical areas. The idea is to mix the two models. The study that we propose will consider the problems of rating evolution and the interdependence among sectors in the same and/or in different countries.

To study this phenomenon we will consider a generalization of the multivariate Markov model that could simultaneously take into account the dynamic evolution of credit risk ratings and the regional interdependence Input-Output model.

Asymptotic Behavior of the Processes Describing Some Insurance Models

Ekaterina Bulinskaya

Department of Mathematics and Mechanics, Lomonosov Moscow State University, Russia

We study a class of discrete-time processes arising in various applications such as insurance, inventory and queuing theory and others. They describe the state of input-output models. Output is supposed to be a sequence of i.i.d. random variables. Our aim is to choose the input (control) minimizing a functional on trajectories of the above mentioned processes. Two types of functionals are considered. The first one takes into account the costs associated with system performance, whereas the second one evaluates the ruin probability and the process state just before the ruin and at the ruin time.

At first we obtain the form of optimal control under assumption that all the parameters and underlying processes distributions are known. After that the sensitivity analysis is carried out, see, e.g. [1].

Next step is investigation of the incomplete information case, see, e.g., [2].

We consider two possibilities, namely, the distribution form is known up to parameters or it is completely unknown. In the former case we introduce various parameters estimates and study their impact on the control. In the latter case we introduce the notion of empirical asymptotically optimal control and study its form.

The last step is the proof of limit theorems for appropriately normalized continuous-time processes based on discrete-time ones describing the systems under consideration. Thus, it could be possible to use the diffusion approximation for evaluation of the system performance quality.

Keywords: Incomplete information, empirical asymptotically optimal policies, limit theorems

Acknowledgement. The research is partially supported by RFBR grant 13-01-00653.

References

1. E.Bulinskaya. Sensitivity analysis of some applied models. *Pliska Studia Mathematica Bulgarica*, 18, 57-90, 2007.
2. E.Bulinskaya. Some aspects of decision making under uncertainty. *Statistical Planning and Inference*, 137, 8, 2613-2632, 2007.

Optimal exponential inequalities for the distribution tails of multiple stochastic integrals with Gaussian integrators

Alexander Bystrov

Department of Mechanics and Mathematics, Novosibirsk State University, Russia

We study multiple stochastic integrals of nonrandom function with respect to product-measure generated by Gaussian process on $[0,1]$. Such integrals (with different Gaussian integrators) arise, for example, as weak limits for sequences of so-called V-statistics (von Mises statistics) based on samples from independent or weakly dependent observations. We describe wide class of Gaussian integrators, for which we obtain upper exponential bounds for distribution tails of such integrals. Integral kernels suppose to be bounded. We also discuss the optimality of these inequalities.

Keywords: Stochastic integrals, exponential inequalities, weak dependence, V-statistics.

A fuzzy regression model for non-convex fuzzy numbers: the crisp input - fuzzy output case

Antonio Calcagn, Luigi Lombardi

Department of Psychology and Cognitive Science, University of Trento, Italy

In this contribution we propose a least square approach to fuzzy linear regression analysis for non-convex data which generalizes a previous proposal for convex data (see: D'Urso and Gastaldi[1], Coppi et al.[4]). A formal definition of non-convex fuzzy number is provided and a fuzzy linear regression model for this type of fuzzy data is introduced. We also applied our fuzzy regression model to a simulated dataset. Finally, some relevant comments about limitations, potential new applications, and extensions of this approach are presented.

Keywords: Fuzzy data analysis, fuzzy regression model, non-convex fuzzy data.

Stochastic models for the chemostat at different scales

Fabien Campillo¹, Marc Joannides^{1,2}, Irène Larramendy Valverde²

¹Project-Team MERE, INRIA/INRA, UMR MISTEA, ²Université Montpellier 2, France

The evolution of the state of a single species/single substrate chemostat is usually described by a set of ordinary differential equations (ODE) derived from a mass balance principle. In this case, the modeling process relies on the fact that the stochastic effects can be neglected thanks to the law of large numbers. This is possible only at macroscopic scale, for high population sizes, and under homogeneity conditions. At all other scales or when the homogeneity conditions are not met, random effects cannot be neglected. Our goal is to establish a set of stochastic models that are valid at different scales: from the small population scale to the scale immediately preceding the one corresponding to the deterministic model. At a microscopic scale we present a pure jump stochastic model that gives rise, at the macroscopic scale, to the ordinary differential equation model. At an intermediate scale, an approximation diffusion allows us to propose a model in the form of a system of stochastic differential equations. The convergence of the pure jump model or of the diffusion approximation to the deterministic model can be rigorously established.

We expound the mechanism to switch from one model to another, together with the associated simulation procedures. Three associated simulation algorithms that will be valid at different scales are presented. We also describe the domain of validity of the different models. The pure jump model can be exactly simulated thanks to the Gillespie algorithm, also called stochastic simulation algorithm. In standard cases, that is for high population levels, this procedure is not feasible as it requires us to simulate too many events. In this case, we present the Poisson approximation and the normal approximation, both in discrete-time.

GENERAL DIAMETER-HEIGHT EQUATIONS FOR *Ochroma lagopus* S.W., COAST REGION, ECUADOR

Álvaro Cañadas¹, František Vilčko², Diana Rade¹, Carlos Zambrano³

¹Universidad Técnica del Norte, Facultad de Ingeniería en Ciencias Agropecuarias y Ambientales, Ecuador, ²University of Göttingen, Institute of Forest Management, Germany, ³Instituto Autónomo de Investigaciones Agropecuarias (INIAP), Ecuador

The Balsa exports of the Ecuador have registered an increment of 560% among the 2007-2012. Nevertheless, deficient information about Balsa plantation exists. The object of the present investigation was to develop and compare generalized diameter-height equations for pure stand of

Balsa (*Ochroma lagopus* Sw.). The data used to fit the equations consisted of 920 diameter-height observations samples from 120 plots installed in pure, uniformly stocked stands covering a wide range of forest stand and site conditions. In a first step, nonlinear least square regression was used to test the performance of two published equations; one derived from Richards (equation 1) and showed a 91% of efficiency and the other one from Schumacher function (equation 2) with an efficiency of 86%. In both cases, the residuals indicated no severe violation of the fundamental statistical assumptions. In a second step, equation (2) was modified resulting a new equation which showed only a marginal improvement. Finally, the efficiency was evaluated for all equation using an independent data set. The strongest correlation between observed and predicted values was obtained with equation (1). Equations (1), (2) and (3) could be used predict the individual Balsa tree in the province of Los Ríos.

Keywords: Balsa, Dominant Height, Nonlinear regression, Richards function, Schumacher function

A comparison of artificial neural networks to forecast nitrogen dioxide concentrations

Carmen Capilla

Department of Applied Statistics and Operations Research and Quality,
Polytechnic University of Valencia, Spain

The objective of this work is to investigate the forecasting capability of several neural network models. A time series of nitrogen dioxide (NO₂) concentrations is used as an example. One of the important air pollution indicators is NO₂. How to monitor and predict air quality indicators has become a problem in environmental management. The main methods used for air quality forecasting are based on deterministic models and statistical methods. Deterministic models aim to reproduce physical processes, but they do not often give satisfactory results. Neural networks have been shown to be effective alternatives to these models and to more traditional statistical techniques, when analyzing chaotic and non-linear relationships. Especially, the "multi-layer perceptron (MLP)" has been applied in air quality prediction in the last decade. In this work, a comparison of artificial neural network models is presented to predict hourly NO₂ concentrations from local pollutants concentrations, traffic, meteorological data, and periodic components. Three different methods are compared: feed-forward MLP networks, radial basis function (RBF) networks and generalized regression (GR) networks. The study area is in Valencia (Spain) with hourly data from 2003 to 2005. For monitoring and control purposes an automatic air pollution network is operated in the whole urban area of this city by the local government. Hourly

pollutant, meteorological and traffic observations are available in one background monitoring stations. The meteorological variables are: wind direction, wind speed, temperature, relative humidity, pressure and solar radiation. The traffic data are the hourly number of vehicles. The number of neurons in the hidden layer of the MLP networks is the optimum found by experimentation. The transfer functions selected for the MLP networks are the sigmoid and the hyperbolic tangent for the hidden layer, and linear for the output layer. The MLP model is applied using the Levenberg-Marquard and the scaled conjugate gradient learning algorithms. The spread parameter for the RBF and GR networks are also found by experimentation. The output of the models is NO₂ prediction 24 hours in advance. The inputs are nitrogen monoxide (NO) concentrations, traffic and meteorological data, and periodic components (sine and cosine terms for the daily and weekly cycles). The models are trained on data from 2003. Data from 2004 are used as the validation set and observations from 2005 are the test data set. The model evaluation criteria are the mean absolute error, the root mean square error, the mean absolute percentage error and the correlation coefficient (r) between the pollutant observations and the predictions. Bootstrap estimates of standard errors of these criteria are calculated. These estimates were obtained by randomly resampling the test data 1000 times, with replacement. The radial basis approach provides the most accurate forecasts.

Keywords: Neural networks, Bootstrap, Air quality prediction.

A Novel Pointer-Encoding Genetic Programming Algorithm for Classification Problems

L. Carro-Calvo, S. Salcedo-Sanz, A. Portilla-Figueras, S. Jiménez-Fernández, L. Cuadra, E. Alexandre-Cortizo

Department of Signal Theory and Communications, Escuela Politécnica Superior, Universidad de Alcalá, Spain

This paper presents a novel Genetic Programming (GP) algorithm to solve classification problems. The proposed GP includes a novel encoding of trees in the form of pointers, defined over a two level matrix structure. The advantages of this novel encoding is that it allows defining trees of any deep, in a simple way, and it is robust and compact, perfect to apply crossover operators in a straightforward manner. The proposed encoding also allows a very fast evaluation of subtrees, which is of interests in different applications. We have tested this novel GP in several classification problems, obtaining important advantages versus the classical GP and different classes of neural networks.

Some optimization and decision problems in proportional reinsurance

Castañer, A., Claramunt, M.M., Mármol, M.

Universitat de Barcelona, Spain

In this paper, the classical risk theory model is modified with the introduction of a threshold proportional reinsurance strategy (Castañer et al. (2012)). This dynamic strategy considers a retention level that is not constant and depends on the level of the surplus with respect to a threshold. Then we obtain two different values for the premium income retained by the insurer. We consider that the retention levels give new positive security loadings for the insurer, i.e. the net profit condition is always fulfilled.

In this model with threshold proportional reinsurance strategy the discounted penalty function (Gerber and Shiu (1998)) behaves differently, depending on whether initial surplus is below or above the threshold level. This function can be used to study the traditional quantities of interest in classical ruin theory, such as ruin probability, time of ruin or deficit at ruin. Therefore, depending on the penalty function, we can obtain different interpretations for the Gerber-Shiu function.

Our analysis is focused on the distribution function and the moments or the deficit at ruin if ruin occurs, using the Gerber-Shiu function, assuming exponential interoccurrence times and individual claim amount phase-type. So we will consider only a certain subgroup of penalty functions.

The theoretical results are used to examine several optimization and decision problems of the insurer with respect to the reinsurance strategy, using the probability of ruin, the Value at Risk and Tail Value at Risk of the deficit at ruin if ruin occurs as risk measures.

Keywords: Deficit at ruin, Gerber-Shiu function, Risk measures

Regular E-optimal Spring Balance Weighing Designs With Correlated Errors

Bronisław Ceranka, Małgorzata Graczyk

Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poland

In the paper, the problems linked with the theory of determining unknown measurements in the spring balance weighing design are discussed. We assume that the result of an experiment we can describe as linear combination of unknown measurements of objects with factors of this combination equal to 0 or 1. Additionally, we assume that the measurements errors are equal correlated and they have the same variances.

The topic is focus on the determining the maximal eigenvalue in the design. The conditions that determine optimal design, relations between the parameters and the forms of matrix of regular E-optimal design are presented. Moreover, the constructing method of the E-optimal design, based on the incidence matrices of balanced incomplete block designs, is given.

Keywords: balanced incomplete block design, E-optimal design, spring balance weighing design.

Relations Between Regular A-Optimal Chemical Balance Weighing Designs with Diagonal Covariance Matrix of Errors

Bronislaw Ceranka, Malgorzata Graczyk

Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poland

In the paper, we study the problem of estimation of measurements of objects using a weighing design. We assume that the measurements errors are uncorrelated and they have different variances. We consider two kinds of weighing designs: chemical balance weighing design and spring balance weighing design. Moreover we consider such designs under criterion of A-optimality.

The necessary and sufficient conditions for regular A-optimal biased spring balance weighing designs with the diagonal covariance matrix of errors and regular A-optimal chemical balance weighing designs with diagonal covariance matrix of errors are given and the relations between those designs are investigated.

Keywords: A-optimal design, chemical balanced weighing design, spring balance weighing design.

Bandwidth selection for mean shift clustering

José E. Chacón¹, Tarn Duong²

¹Departamento de Matemáticas, Universidad de Extremadura, Spain,

²Laboratoire de Statistique Theorique et Appliquée (LSTA), Université de Paris 6, France

The mean shift algorithm was introduced in Fukunaga and Hostetler [1]. It is an iterative procedure which, at every step, shifts the point obtained in the previous iteration in the direction of the estimated normalized density gradient, producing a convergent sequence that transports the initial value to a local maximum of the density estimate along the steepest ascent path. This algorithm induces a partition of the data in a natural way, by assigning the same cluster to all the data points that lead

to the same local maximum when the convergence of the iterative procedure is reached. Notice that this methodology does not require the number of clusters to be specified in advance, and that it allows clusters of arbitrary shape to be discovered. In this communication we study the properties of this algorithm when kernel methods are used to estimate the normalized density gradient, and propose several data-driven bandwidth selectors to make this methodology fully automatic. The procedure is illustrated with simulations and real data examples.

Keywords: bandwidth selection, mean shift algorithm, nonparametric clustering.

References

[1] Fukunaga, K. and Hostetler, L.D. (1975) The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE T. Inform. Theory*, **21**, 32 – 40.

Finding the best combination of variables with respect to an ROC curve type measure under a continuous reference

Yuan-chin Ivan Chang

Institute of Statistical Science, Academia Sinica, Taiwan

As the classification problems become a popular task in many statistical applications, how to evaluate the performance of classifier becomes an interesting and important issue. The receiver operating characteristic (ROC) curve is a useful tool for such a purpose. However, to apply this method and the methods derived from it, there must be a binary label for each subject. However, in many applications, such as medical diagnosis, this kind of binary label may not be naturally available, or is dichotomized from a continuous measurement based on the current knowledge about the classification problem. This happens situation happens especially in biological/medical areas, where the cutting point of gold standard varies as advances are made in diseases. When the reference variable is continuous and there is no confirmative threshold to transform it into binary, ROC curve becomes less useful. In Obuschowski (2006), a novel measure, which does not rely on the cutting point of the gold standard, is proposed, which extends the AUC index along the nonparametric interpretation of area under the curve. Here a more general extension of this type measure is discussed, which not only maintains the nonparametric interpretation, but also the geometric nature of the area under curve. In addition, the measure proposed in Obuchowski (2006) can be treated as a special case of this novel measure. Algorithms for finding the best linear combination of variables with respect to this measure are studied that can be applied to the case when number of variables outnumbers the sample size. The asymptotic

properties of the corresponding estimates are studied, and their performance is illustrated using both synthesized and real data sets.

Keywords: ROC curve, area under curve, diagnosis

Scan Statistics for Monitoring Data Modeled by a Negative Binomial Distribution

Jie Chen¹, Joseph Glaz²

¹University of Massachusetts Boston, ²University of Connecticut, USA

In this article we investigate the performance of approximations and inequalities for the distribution of scan statistics for independent and identically distributed observations from geometric and negative binomial distributions. The use of scan statistics is discussed for prospective and retrospective type experiments.

These scan statistics can be also used in a sequential type experiments for monitoring data, modeled by a geometric or a negative binomial distribution, for detecting a local change in the waiting time for a specified event (batch of events). Potential applications include: business, ecology, criminology, entomology, quality control and sampling schemes. Numerical results are presented to evaluate the performance of the approximations discussed in this article. Extensions to variable window scan statistics, multiple scan statistics and two dimensional data are mentioned as well.

Keywords: Approximations, Inequalities, Moving sum, Sequential monitoring, Waiting time.

SOME SEQUENTIAL MONTE CARLO TECHNIQUES FOR DATA ASSIMILATION IN A PLANT GROWTH MODEL

Yuting Chen, Samis Trevezas, Paul-Henry-Cournède

EPI DIGIPLANTE, INRIA Saclay Île-de-France, France
Laboratory of Applied Mathematics and Systems, Ecole Centrale Paris,
France

Data assimilation techniques have received considerable attention due to their capability to improve prediction and the most important applications concern weather forecasting and hydrology. Among many competing data assimilation approaches, those based on sequential Monte Carlo (SMC) methods, known as “particle filters”, have gained their popularity because they are adaptive to nonlinearity and non-Gaussianity. In this study we test the performance of two SMC methods to predict biomass production and allocation in a dynamically evolving plant-growth model (GreenLab model, see Cournède *et al.*[1]) that can

be formalized as a nonlinear state space model. The first method concerns a post-regularized particle filter (post-RPF, see Musso *et al.*[2]) which uses a mixture of Gaussian kernels (or more generally a kernel based method) to avoid sample impoverishment in the resampling step and the second method involves a convolution particle filter (Campillo and Rossi[3]) which regularizes as well the densities of the observation noise. Both approximate Bayesian estimation techniques deal simultaneously in their state vector fixed model parameters and state variables. We show that these methods perform well in realistic scenarios with sparse observations and discuss their limitations. Outside the context of data assimilation, we also present a maximum likelihood approach based on a stochastic version of an EM (Expectation-Maximization) algorithm, where the E-step can be approximated by the aforementioned SMC methods and discuss the pros and cons of the resulting algorithm. The performance of our methods is illustrated on data from the sugar-beet.

Keywords: plant growth model, data assimilation, sequential Monte-Carlo methods, kernel based method, stochastic EM algorithms, sugar-beet.

References

1. Cournède P.-H., Kang M. Z., Mathieu A., Barczi, J.F., Yan, H.P., Hu, B.G. and De Reffye, P. Structural Factorization of Plants to Compute their Functional and Architectural Growth. *Simulation* 82(7), 427-438, 2006.
2. Musso C., Oudjane N. and Legland F. Improving regularized particle filters. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, 12, 247–271. Statistics for Engineering and Information Science, 2001.
3. Campillo, F. and Rossi, V. Convolution Particle Filter for Parameter Estimation in General State-Space Models. *IEEE Trans. Aero. Elec. Sys.* 45(3), 1063-1072, 2009.

Beran's nonparametric estimator of a regression reliability function and its application for goodness-of-fit testing

Ekaterina Chimitova, Victor Demin, Vladislav Schekoldin

Novosibirsk State Technical University, Russia

One of the most popular nonparametric estimators for a conditional reliability function under the given value of the observed covariate is the Beran estimator. In this paper we propose the method of choosing an optimal smoothing parameter for the Beran estimator, which is based on the minimization of the observed lifetime deviation from a nonparametric estimator of the inversed reliability function, obtained by kernel smoothing. By means of the Monte-Carlo simulations it has been shown

that the method usage results in more precise estimates than using a fixed smoothing parameter. We propose a goodness-of-fit test for parametric and semiparametric reliability regression models, which is based on a distance between the Beran nonparametric estimator and the tested conditional reliability function for given values of the covariate. The power of the proposed test has been investigated for various pairs of competing hypotheses.

Keywords: Conditional reliability function, Beran's estimator, kernel smoothing, goodness-of-fit tests.

On testing goodness-of-fit with the semi parametric Accelerated Failure Time model

Ekaterina Chimitova, Natalya Galanova

Department of Applied Mathematics and Computer Science, Novosibirsk State Technical University, Russia

This paper is devoted to the problems of semi parametric Accelerated Failure Time Model construction. Maximum likelihood estimates of the regression parameters and its properties are considered for both censored and complete data. Various parameterizations of covariate function are applied to the model construction: log-linear model, Arrhenius model, power rule model. Nonparametric estimates of the survival functions are investigated in the paper. Constant and time-dependent covariates are used in the modeling procedures. The extrapolation ability of the semi parametric AFT model for covariate values which are not presented in the plan of the experiment is discussed. Testing goodness-of-fit procedures for semi parametric Accelerated Failure Time Model are presented in the paper for both complete and censored data. The comparison of the parametric and semi parametric approaches to the construction of the Accelerated failure Time model is made for real data of insects' survival.

Keywords: semi-parametric Accelerated Failure Time Model, censored data, maximum likelihood, residuals, time-dependent covariates, survival function, log-linear model, Arrhenius model, power rule model, goodness-of-fit testing, insects' survival data

Analysis of Multivariate Functional Data

Jeng-Min Chiou

Institute of Statistical Science, Academia Sinica, Taiwan

Multivariate functional data are increasingly common owing to advances in modern technology. They comprise multiple functional measurements

recorded simultaneously in contrast to a single functional measurement for univariate functional data. We consider a regression model where both the response and the predictor variables contain multiple random functions. We propose a multivariate functional principal component approach that considers varying extent of variation between the random functions and takes advantage of component dependency through the pairwise cross-covariance functions. This multivariate approach serves as a basic tool in dimension reduction and analysis of multivariate functional data. We investigate the consistency properties for the estimates of the model components, and derive the asymptotic results for statistical inferences. Based on this multivariate functional principal component approach, we derive a multivariate functional regression model for analysis and prediction of multivariate functional responses. The methods are illustrated using an application with interest relating to traffic flow prediction, where we predict the future traffic flow parameters for a set of up-to-date and partially observed traffic flow trajectories.

Keywords: Confidence Band, Functional Linear Regression, Functional Principal Component Analysis.

Non-Arbitrage Under Additional Uncertainties/Information

Tahir Choulli, Anna Aksamit, Jun Deng, Monique Jeanblanc

Mathematical and Statistical Sciences Department, University of Alberta,
CANADA

In this talk, I will address the question of how an arbitrage-free semimartingale model is affected when this model is stopped at a random horizon or when a honest time is totally incorporated. Precisely, I will focus on weak form of non-arbitrage that is crucial for the existence of optimal portfolio. This type of non-arbitrage is called the No-unbounded-Profit-with-Bounded-Risk (called NUPBR hereafter) concept. It is also known --in the literature-- as the first kind of non-arbitrage. I will present many new results that can be classified into two principal types of results. First, I will provide necessary and sufficient conditions on the random time (default time/exit time/random horizon) such that the Non-arbitrage concept remains valid for any model stopped at this random time. The second type of results consists of considering a fixed (but arbitrary) semimartingale and a random time, and derive the necessary and sufficient conditions under which this semimartingale stopped at this random time still fulfill the non-arbitrage concept. When a class of honest times into the model will be incorporated into the model, we also provide necessary and sufficient conditions that guarantee the preservation of the non-arbitrage concept. Our class of honest times is much larger than the class of all stopping times. More importantly, our condition that defines our class of honest time sounds play important role in classifying honest times. The crucial stochastic tool that drives our analysis lies in

the {it optional stochastic integral} (or {it compensated stochastic integral}) and its extension. This talk is based on a joint paper with Anna Aksamit, Jun Deng, and Monique Jeanblanc.

Keywords: Non-arbitrage, Defaultable Market, Random Horizon, Honest Time, Stochastic Calculus, Martingale theory

Prediction for Count Time Series

Vasiliki Christou, Konstantinos Fokianos

University of Cyprus, Department of Mathematics and Statistics, Cyprus

We study inference and diagnostics for count time series regression models which include a feedback mechanism. In particular, we are interested in negative binomial processes for count time series. We study probabilistic properties and quasi likelihood estimation for this class of processes. The resulting estimators are consistent and asymptotically normally distributed. The key observation in developing the theory is a mean parameterized form of the negative binomial distribution. In addition, we provide probabilistic forecasts based on the assumption of negative binomial or Poisson distribution and we propose the use of probability integral transformation histogram (PIT), marginal calibration plot and scoring rules to assess the predictive performance and rank the competing forecast models.

Keywords: Diagnostics, Prediction, Quasi-maximum likelihood estimation.

The Building of Credit Scoring System on the Residential Mortgage Finance by Logistic Regression and Classification Tree_ A Case Study From Taiwan

Rwei-Ju Chuang

Fu-Jen University, Taiwan

The volume of credit business in the residential mortgage finance has greatly expanded and the use of credit scoring through the evaluation of large credit portfolio becomes crucial to guard against any management risk. The objective of this study is to devise a credit scoring system for finance granting decisions. We describe statistical method to create scorecards and show how the result of the model is applied to calculate score point weights. Scorecards are built using the logistic regression method which estimates the relationship between the individual characteristics and the log of the odds (risk) so that the score point weights can be calculated directly from the regression coefficients. The model performance is monitored by the model validation and

classification error. We propose an alternative measure for power of model discriminations and credit-granting decisions.

Keywords: Credit Scoring, Logistic Regression Model, Classification Tree.

An Economic Order Quantity Inventory Model with Time Dependent Weibull Deterioration and Trended Demand

Chukwu W.I.E, Sanni S.S

Department of Statistics, University of Nigeria Nsukka, Nigeria

A single-item economic order quantity model is presented in which inventory is depleted mainly due to demand and partly due to deterioration. The rate of deterioration is taken to be time dependent, and the time to deterioration is assumed to follow three-parameter Weibull distribution, the demand rate is quadratic function and shortages are allowed in the inventory and are completely backlogged. The Weibull instantaneous rate function describes different situations of deterioration while the quadratic demand function depicts the various phases of market demand. We provide simple analytical tractable procedures for deriving the model and also establish the necessary and sufficient conditions for the optimal replenishment policy for the inventory model. Numerical examples are given to illustrate the solution procedure and sensitivity analysis is conducted to evaluate the responsiveness of the proposed model to changes in the model parameters

A strong optimality result for anisotropic self-similar textures

M. Clausel¹, B. Vedel²

¹Laboratoire Jean Kuntzmann, Université de Grenoble, ²Laboratoire de Mathématiques et Applications des Mathématiques, Université de Bretagne Sud, Université Européenne de Bretagne Centre Yves Coppens, France

In [13,14], we proposed a method to characterize jointly self-similarity and anisotropy properties of a large class of self-similar Gaussian random fields. We provide here a mathematical analysis of our approach, proving that the sharpest way of measuring smoothness is related to these anisotropies and thus to the geometry of these fields.

Keywords: Operator scaling Gaussian random field, anisotropy, sample paths properties, anisotropic Besov spaces.

The SCR Adequacy according to the Volatility Longevity Shocks

Mariarosaria Coppola¹, Valeria D'Amato²

¹ Department of Political Sciences, University of Naples Federico II

² Department of Statistics and Economics University of Salerno

The improvements in longevity observed in many countries over the past century have been significant. The risk that the longevity experience is higher than the one forecasted, i.e. longevity risk, is explicitly considered in Solvency II standard formula as a sub-module of the life underwriting risk module. The life underwriting risk module includes all the life insurance and reinsurance obligations, except the SLT health insurance obligations (EIOPA 2012), where the longevity risk is one of the seven sub-modules. According to Solvency II, solvency capital requirements (from herein SCRs) can be computed by a standard formula or an internal model. Nevertheless, the scenario in which the insurance companies operate is often more complex than that one assumed by the standard formula. According to the standard formula the SCR is represented as the change in net asset value due to longevity shock which is a permanent 20% reduction of mortality rates for all ages. A constant shock is not reasonable for all ages and maturities. The scenario related to the standard formula may lead to a biased allocation of capital, because of the volatility of longevity phenomenon in respect of different ages. In this paper we examine the adequacy of SCRs on the basis of the standard formula. To correctly calculate the solvency capital requirement we follow a multi-period approach in the sense that we evaluate at the beginning of each year the amount of capital that the insurer need to meet its future obligations year by year till the contract will be in force. We examine the adequacy of the shocks structure suggested by the standard formula studying its impact on the SCR for longevity risk (SCRLong) and liabilities at different ages.

Keywords: Solvency II, Solvency Capital Requirement, Longevity shocks, Multi-period approach

The Impact of a Pandemic Influenza on Mortality, Temporary Disability and Hospitalization Risks for Protection Insurance, and Hedging Strategies

Marine Corlosquet-Habart^{1,2,3}, Jacques Janssen⁴, Philippe Lenca², Raimondo Manca⁵

¹EURIA, Université de Bretagne Occidentale, France, ²Telecom Bretagne, Lab-STICC, France, ³BNP Paribas Cardif, France, ⁴Solvay Brussels School for Economics and Management, Université Libre de Bruxelles, Belgium, ⁵"La Sapienza" University, Dipartimento Di Metodi E Modelli Per L'Economia, Italia

As far as Protection Insurance is concerned, the emergence of an influenza pandemic could lead to an excess of claims in the insured population. A pandemic would obviously have a significant impact on

Death risk, but could also affect the risk of Temporary Disability and Hospitalization. The aim of this article is to determine how to model the effects of a pandemic event on mortality risk, and to analyze how a pandemic may affect the risk of Temporary Disability and Hospitalization. Knowing that a pandemic could cause bankruptcy in the worst scenarios, an insurance company should consider buying adequate coverage. The last part of this article is devoted to the search for a solution to hedge against a pandemic event. This overview is not limited to standardized products sold by reinsurers, but more elaborate and exotic offers are described, with their advantages and drawbacks.

Keywords: Pandemic Risk, Influenza, Protection Insurance, Mortality Risk, Temporary Disability, Hospitalization, Hedging Strategies, Reinsurance, Securitization

Assessing the importance of risk factors in distance-based generalized linear models

Teresa Costa, Eva Boj, Josep Fortiana, Anna Esteve

Facultat d'Economia i Empresa, Universitat de Barcelona, Spain

In distance-based generalized linear models (DB-GLM) (Boj et al. 2012) there is an additional non-linearity layer between observed predictors and response, due to the presence of latent variables derived from the distance function. This indirect relationship prevents us from interpreting linear predictor coefficients as is done in ordinary GLM.

It is of interest to quantify the relative leverage of risk factors when we apply DB-GLM in actuarial problems as are a priori rate making or claim reserving. With this aim we study some local -valid in a neighborhood of a given point in predictor space- measures of the relative importance of each observed predictor and define a version of the F-test statistic for selecting explanatory variables, whose p-values can be estimated by bootstrap.

This research is part of the project: Semiparametric and distance-based methodologies with applications in bioinformatics, finance and risk management (grant MTM2010-17323).

Keywords: Distance-based generalized linear model; predictor influence; actuarial science.

Reference: Boj, E., Delicado, P., Fortiana, J., Esteve, A. and A. Caballé (2012). Local distance-based generalized linear models using the dbstats package for R. Documentos de Trabajo de la Xarxa de Referència en Economia Aplicada (XREAP), XREAP2012-11.

Undergraduate Students' Career in Italy: a non Homogeneous Markov Approach with Fuzzy States

Franca Crippa¹, Marcella Mazzoleni², Mariangela Zenga²

¹Department of Psychology, Università di Milano-Bicocca, ²Department of Statistics and Quantitative Methods, Università di Milano-Bicocca, Italy

The Italian university administrative organisation does not allow to relate students career accumulation of credits to their academic years of registration clearly, as it is instead the case in the majority of the European Higher Education (HE) systems. For instance, two freshmen might belong to different progression levels according to the actual number of exams they passed; nonetheless, regardless of the amount of exams they succeeded in the two past syllabuses, both these two freshmen at the end of their first academic year are equally permitted to register at their second academic year and therefore show an identical administrative situation. With reference to the aforementioned HE system, in this paper we apply the theory of non homogeneous Markov systems with fuzzy states, in order to describe the actual students' educational progress in two different faculties of the university of Milano-Bicocca. The levels of students career progression will be related to the academic years using a stochastic model that assumes the progress levels as fuzzy states [Symeonaki, Kalamatianou, 2011] with a membership function related both to the exam's final grade and to the time needed to pass the exam in comparison with the schedule provided in the syllabus. Moreover, the membership function that relates each fuzzy state to the administrative situation will be analysed, so as to derive further insight in students' progression, with particular focus on the possibility of obtaining a function of undergraduates' difficulty in passing the various exams.

Keywords: Progression of students carriers, Non Homogeneous Markov system, fuzzy states, membership function, exams difficulty.

How high is it possible to extend the retirement age? Life expectancy vs. healthy life expectancy

Kornélia Cséfalvaiová, Martina Šimková, Jitka Langhamrová

The University of Economics, The Faculty of Informatics and Statistics,
Department of Demography, Czech Republic

The European Commission has declared 2012 The European year for active aging and solidarity between generations. This issue calls each society to analyze the process of population aging and its social and economic consequences. Currently active aging can be the key to solve the demographic trends, because Europe is the continent with the

slowest population growth. Current demographic trends are connected with low fertility rate, increasing number of seniors and increasing life expectancy. These unfavorable demographic changes have a socio-economic impact on society and this situation calls each society to deal with quality of life of older people and with their economic activity in the labor market. Submitted study follows the question of increasing the retirement age as a possible way for growing economic activity of the post-productive population. We ask whether increasing the retirement age is beneficial or not. Extending life expectancy may have a positive effect on the productivity of seniors and their life satisfaction in case seniors live and work in good health without diseases. The aim of this work is to characterize the evolution of population aging and compare life expectancy and healthy life expectancy in the Czech Republic and in selected European countries. Another aim is reflecting the retirement age and highlighting its likely future development – how far is it possible to extend the retirement age?

Keywords: aging of population, retirement age, healthy life expectancy

Multiple Population Projections by Lee Carter Models

Valeria D'Amato¹, Steven Haberman², Gabriella Piscopo³, Maria Russolillo¹, Lorenzo Trapani²

¹Department of Economics and Statistics, University of Salerno, Italy,

²Faculty of Actuarial Science and Insurance, Cass Business School, City University, UK, ³Department of Mathematics for Decisions, University of Florence, Italy

The academic literature in longevity field has recently focused on models for detecting multiple population trends ([9],[17],[20], etc.). In particular increasing interest has been shown about “related” population dynamics or “parent” populations characterized by similar socio-economic conditions and eventually also by geographical proximity. These studies suggest dependence across multiple populations and common long run relationships between countries (for instance see [13]). In order to investigate cross-country longevity common trends, we adopt a multiple population approach. The algorithm we propose retains the parametric structure of the Lee Carter model, extending the basic framework to include some cross dependence in the error term. As far as time dependence is concerned, we allow for all idiosyncratic components (both in the common stochastic trend and in the error term) to follow a linear process, thus considering a highly flexible specification for the serial dependence structure of our data. We also relax the assumption of normality, which is typical of early studies on mortality [14] and on factor models (see e.g. the textbook by [1]). The empirical results show that the Multiple Lee Carter Approach works well in presence of dependence.

Keywords: Serial and Cross-sectional Correlation, Factor Models, Vector Auto-Regression, Sieve Bootstrap, Lee Carter model

A semi-Markovian Present Worth Analysis and its applications

Guglielmo D'Amico

Department of Pharmacy, Università "G. d'Annunzio" di Chieti-Pescara, Italy

In this paper, a general approach is provided to the classical present worth analysis. The model is based on finite semi-Markov chains that model the growth rate of the considered financial process. The financial process can be the dividend paid by a stock or more in general whatever kind of cash-flow. Applications of the model in the valuation of stocks and in the valuation of real options are discussed.

Keywords: fundamental analysis, reward process, difference equations.

Some Remarks on Income Inequality Measurement

Guglielmo D'Amico¹, Giuseppe Di Biase¹ Raimondo Manca²

¹ Department of Pharmacy, Università "G. d'Annunzio" di Chieti-Pescara, Italy, ² Dipartimento di Metodi e modelli per l'Economia, il Territorio e la Finanza, Università di Roma "La Sapienza", ITALY

In this study we model the income dynamic by means of a stochastic model. We classify the economic agents into K mutually exclusive groups depending on the individual income. Since the income evolves in time, we use Markov and semi-Markov models to reproduce the evolution of the population among these states. We propose dynamic indices to measure the income inequality in the population. We discuss the case in which each agent has a deterministic income depending on the temporary own class. Furthermore, we extend the analysis to cover the case of a random income by means of reward processes. We investigate the possibility of the adoption of stochastic order relations on the model and the consequences on the dynamic inequality indices. Finally we show practical applications of the model and we pay particular attention to the effects of the fiscal policy on the income inequality.

Keywords: income inequality, dynamic indices, semi-Markov

Multivariate Weighted Indexed Semi-Markov Models in Stock Market

Guglielmo D'Amico¹, Filippo Petroni²

¹ Department of Pharmacy, Università "G. d'Annunzio" di Chieti-Pescara, Italy, ² Dipartimento di Scienze Economiche e Aziendali, Università degli studi di Cagliari, Italy

In this study we model the high frequency price dynamics of traded stocks by using an indexed semi-Markov approach. More precisely we assume that the intraday returns are described by a discrete time homogeneous semi-Markov model which depends also on a memory index. The index is introduced to take into account periods of high and low volatility in the market. The model is extended to take into account also the correlation between stocks by adding a dependence structure in the transition probability matrix. First of all we derive the equations governing the process and then theoretical results are compared with empirical findings from real data. In particular we analyzed high frequency data from Italian and German stock market from first of January 2007 until end of December 2010. We show that the model is able to reproduce empirical findings.

Keywords: high frequency data, autocorrelation, cross-correlation.

New Results for an Evolution Problem

Zoubir DAHMANI¹, Soumia BELARBI²

¹Laboratory LPAM, Faculty SEI, UMAB, University of Mostaganem,

²Faculty of Mathematics, USTHB, Algeria

In this talk we discuss new results for the nonlinear evolution equation of order α , ($0 < \alpha \leq 1$). The problem is given by:

$$D^\alpha x(t) = f(t, x(t)); 0 \leq t \leq 1, \quad (1)$$

$$x(0) + x(1) = \int_0^1 g(\tau) d\tau,$$

where D^α denotes the Caputo derivative, and $f: [0, 1] \times E \rightarrow E$ is continuous, such that E is a Banach space and $C([0, 1], E)$ is the Banach space of all continuous functions from $[0, 1] \rightarrow E$ endowed with a topology of uniform convergence with the norm denoted by $\|\cdot\|$.

We give explicit solutions of the above problem:

Theorem 1: A solution of the the fractional evolution problem (1) is given by:

$$\begin{aligned}
 x(t) = & (1-t) \int_0^1 g(\tau) d\tau + t - \frac{t^2}{\Gamma(\alpha+1)} + \frac{t^\alpha}{\Gamma(\alpha+2)} \\
 & - \frac{t}{\Gamma(\alpha+1)} \int_0^1 (1-\tau)^{\alpha+1} f(\tau, x(\tau)) d\tau + J^\alpha f(t, x(t)).
 \end{aligned} \tag{2}$$

Then, using the fixed point theorem and the contraction mapping properties in Banach spaces, we give new sufficient conditions for the existence of a unique solution of the problem (1).

Our third theorem deals with the existence, at least, of a solution of (1). We use a classical theorem to establish sufficient conditions for the proof of this result.

At the end, we give some illustrative examples to verify the accuracy of these evolution results.

Robustness study of the hedging of European claims

C. Daveloose, A. Khedher, M. Vanmaele

Catherine Daveloose

Ghent University, Belgium

Models admitting jumps seem to fit realistic asset prices more properly than continuous models. Although it is appropriate to take jumps into account, they are not easy to handle. Especially processes with infinite activity complicate simulations. Asmussen and Rosinski (2001) launched the idea to approximate a Lévy process by replacing the jump part existing of the small jumps (with size smaller than $c < 1$) by a scaled Brownian motion. The scale is given by the standard deviation of the small jumps. For c tending to zero, this approximation clearly converges in distribution to the original Lévy process. The question rises how this approximation influences the price and hedging strategies of European options on assets modeled by exponential Lévy processes. For the pricing we apply the Fourier approach as described in Eberlein et al. (2010). Because of the presence of jumps, the market is incomplete and there exist many martingale measures. We focus on the Esscher transform, the minimal entropy martingale measure and the minimal martingale measure. Also the delta and quadratic hedging strategies are obtained through a Fourier approach. For the quadratic hedging we consider a martingale and semimartingale setting as in Hubalek et al. (2006). In this paper we prove convergence of the option price, delta and quadratic hedging strategies for c tending to zero. In other words, it is

justified to use the approximation and facilitate numerical experiments with it.

Keywords: Exponential Lévy models, European options, Martingale measures, Pricing, Delta, Quadratic hedging, Robustness

References:

S. Asmussen and J. Rosinski. Approximations of small jump Lévy processes with a view towards simulation. *Journal of Applied Probability*, 38:482-493, 2001.

E. Eberlein, K. Glau, and A. Papapantoleon. Analysis of Fourier transform valuation formulas and applications. *Applied Mathematical Finance*, 17(3):211-240, 2010.

F. Hubalek, J. Kallsen, and L. Krawczyk. Variance-optimal hedging for processes with stationary independent increments. *Annals of Applied Probability*, 16:853-885, 2006.

Random walks with generalized Dirichlet steps

Alessandro De Gregorio

Department of Statistical Sciences, "Sapienza" University of Rome, Italy

We analyze a random walk (also called random flight) moving in the multidimensional real space with a fixed number of changes of direction. The orientations are uniformly distributed on a sphere with radius one. Furthermore the time intervals between two consecutive changes of direction have joint distribution given by a generalized Dirichlet distribution (GD).

For some particular values of the parameters of GD, we are able to obtain the explicit probability distribution of the position reached by the random walk at time $t > 0$.

Keywords: Bessel functions, generalized Dirichlet distributions, random flights, uniform laws.

A parametric non-homogeneous cohort model for the mortality evaluation

**Giovanna De Medici¹, Raimondo Manca¹, Filippo Petroni²,
Giuseppina Ventura¹**

¹MEMOTEF Università di Roma "La Sapienza"

²Università di Cagliari

The mortality forecasting is one of the most important topics of study in the actuarial literature. Indeed, the longevity risk if it is not well

considered in the mortality study can give big problems in the evaluation of liabilities in life and pension insurance models. As it can be easily understood mortality is a non-homogeneous phenomenon and longevity risk is the most important consequence of its non-homogeneity. Furthermore we think that the forecasting of mortality should be done following the evolution of cohorts and not working on cross section data. Taking into account these aspects, the paper presents an algorithm useful for the construction of the mortality evolution of future cohorts. In the last part of the paper some examples constructed by real data will be presented.

Variance reduction techniques for estimating Limited Value-at-Risk and Limited Conditional Tail Expectation measures

Silvia Dedu, Florentin Serban

Department of Applied Mathematics, The Bucharest University of
Economic Studies, Romania

We investigate the performance of using different variance reduction techniques for estimating Limited Value-at-Risk and Limited Conditional Tail Expectation measures. The aggregate loss corresponding to various portfolios composed by financial assets is evaluated, using Limited Value-at-Risk and Limited Conditional Tail Expectation measures. The computational results prove the advantages of using the proposed estimation techniques.

Keywords: Limited Value-at-Risk; Limited Conditional Tail Expectation; Variance reduction techniques; Stop-Loss reinsurance with multiple retention levels; Risk estimation.

Acknowledgement. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-RU-TE-2012-3-0007.

Optimal Portfolio Versus No-arbitrage and its Application

Jun Deng

Mathematical and Statistical Sciences Dept., University of Alberta,
Canada

Since the original work of Arrow and Debreu [1], in the context of discrete market (i.e. finite scenarios and finite number of trading times), it has been proved that the utility maximization admits optimal solution if and only if there exists an equivalent martingale measure (called EMM hereafter) or equivalently there is no arbitrage. This result has been

baptized as the classical fundamental theorem of utility maximization (called FTUM hereafter). However, this theorem {in this current formulation} fails in general continuous-time semi-martingale framework even with smooth enough utility function. In fact, there exist continuous time semimartingale models where the arbitrage opportunities and optimal portfolio both exist. In this paper, we investigate how far we can weaken the non-arbitrage condition as well as the utility maximization problem to preserve their complete and strong relationship described by the FTUM. As the first main contribution, we established the new version of fundamental theorem of utility maximization. Precisely, the equivalence between utility maximization and EMM is reformulated as a localized version by a stationary increasing stopping times and a weaker non-arbitrage concept called No-Unbounded- Profit-with-Bounded-Risk (called NUPBR hereafter). As an important application of our new version of the FTUM, we established the equivalence between the NUPBR condition, the existence of numeraire portfolio, and the existence of solution to the utility maximization under change of equivalent probability measure. The latter fact can be interpreted as a sort of weak form of market's viability. Furthermore, the obtained equivalent probability measure can be chosen as close to the real-world probability measure as we want (but might not be equal).

Keywords: Optimal Portfolio; Utility Maximization; Non-arbitrage; Numeraire Portfolio; Market's Weak Viability; Semimartingale.

This talk is based on the joint work with Choulli T. and Ma J.F.

References

1. Arrow, K. J., Debreu, G.: Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, 265-290 (1954)
2. Artzner, P.: On the numeraire portfolio. *Mathematics of Derivative Securities* 15, 53 (1997)
3. Choulli T., Deng J., Ma J.F.: The fundamental theorem of utility maximization and numeraire portfolio, submitted (2012)
4. Becherer, D.: The numeraire portfolio for unbounded semimartingales. *Finance Stochastics* 5(3), 327-341 (2001)
5. Duffe, D.: *Dynamic asset pricing theory*. Princeton University Press (2008)

A Functional Hodrick-Prescott Filter

Boualem Djehiche, Hiba Nassar

¹Department of Mathematics, The Royal Institute of Technology, ²School of Computer Science, Physics and Mathematics, Linnaeus University, Sweden

We propose a functional version of the Hodrick-Prescott filter for functional data which take values in an infinite dimensional separable Hilbert space. We will further characterize the associated optimal

smoothing parameter when the underlying distribution of the data is Gaussian.

Keywords: Inverse problems, adaptive estimation, Hodrick-Prescott filter, smoothing, trend extraction, Gaussian measure on Hilbert space.

Application of Nonlinear Dynamics to Study the EEG Signals of Subjects in a State of Meditation

Ludmila A. Dmitrieva^{1,2}, Igor E. Kanunikov^{2,3}, Yuri A. Kuperin^{1,2}, Nikolai M. Smetanin¹, Maria A. Shaptiley⁴

¹Department of Physics, Saint-Petersburg State University,

²Department of Liberal Arts and Sciences, Saint-Petersburg State

University, ³Department of Biology, Saint-Petersburg State University,

⁴Scientific Research Center of Health "Exiclub", Russia

In the present paper the data base consisting of 19-channel (international 10/20 system) electroencephalograms (EEG) for 50 subjects recorded in two different states has been studied. The first state (in further - the background) corresponds to the normal state of a subject with closed eyes without movements. The recording of the second state (in further - meditation) was preceded by a 2-minute "instruction" the general meaning of which is that the subject is asked to minimize "the wandering mind," and do not engage in the usual automatic cascade of semantic associations. The main aim of this work was to determine the different quantitative characteristics of the EEG recordings which allow identifying differences between background and meditation. The first group of methods to achieve this goal consisted from the traditional spectral analysis and multifractal analysis of the EEG time series itself. The second group was related to application of nonlinear dynamic methods for study the EEG time series reconstructed attractors. In the spectral analysis, the fast Fourier transform was used to find the power spectra for each channel of the EEG background mode and meditation. It turns out that frequencies corresponding to the maximum of power spectra for subjects in meditation state are less than that in background state. Next, the power spectra of alpha and theta rhythms have been compared. As the result, for subjects in the background mode the power of alpha rhythm is greater than the power of theta rhythm. On the contrary in the mode of meditation the dominance of the theta rhythm has been observed. The similar results have been already discussed by other authors. Among novel approaches of the present paper to study the meditation is multifractal analysis of EEG time series. Construction of multifractal spectra and study of their main characteristics showed the significant difference in the background state and the state of meditation. Another novel approach to study the meditation is application of nonlinear dynamic methods for study the EEG time series reconstructed

attractors. Firstly, the correlation dimensions of reconstructed attractors for EEG signals in background and meditation states have been compared. Secondly, the local divergence exponents (LDR) on the reconstructed attractors have been calculated for background and meditation states. In both cases the significant difference of above quantitative allows to separate the EEG signals in states of meditation and background.

Keywords: electroencephalograms (EEG), meditation, alpha and theta rhythms, multifractal analysis, reconstructed attractors, local divergence exponents (LDR)

Neural Networks Data Mining of EEG Signals Recorded in Different States of Consciousness

Ludmila A. Dmitrieva^{1,2}, Yuri A. Kuperin^{1,2}, Alexey S. Minin³, Maria A. Shaptiley⁴

¹Department of Physics, Saint-Petersburg State University,

²Department of Liberal Arts and Sciences, Saint-Petersburg State University, ³OOO Siemens, Monitoring and Preventive Control, Saint-Petersburg, ⁴Scientific Research Center of Health, OOO "Exiclub", Russia

Usually the process of data mining or extraction of knowledge from the data consists of a number of iterative steps, since the formulation of the purposes depends in some respects on the obtained results. It can include a loop of a feedback that means reformulation the purposes on the basis of the received information. Depending on the purposes and complexity of the data it is possible to use any type of well-known algorithms based on recognition of images, machine learning or the multivariate statistical analysis. The key point here is the detection of originally unknown structures or patterns in the analyzed data. In the present paper we used the neural network intelligent classifier of the data base consisting of 19-channel (international 10/20 system) electroencephalograms (EEG) for 50 subjects recorded in two different states. The first state (in further, the background) corresponds to the normal state of a subject with closed eyes without movements. The recording of the second state (in further, meditation) was preceded by a 2-minute "instruction" the general meaning of which is that the subject is asked to minimize "the wandering mind," and do not engage in the usual automatic cascade of semantic associations. The aim of the present paper is to train the neural network to obtain clusters due to which one could judge on the degree of subject experience in the meditation technique. To do this, the segments of the EEG time series and Fourier spectra (in background mode and meditation) were used to form a training vectors for neural network intelligent classifier. Then we trained a

neural network to classify training vectors into clusters according to their similarity. As a result we obtained the trained neural network which separated the entire data base into a number of clusters. Note that the trained neural network can also be used for on-line evaluation of success of meditative psychological trainings. It is enough to feed to the input of the trained neural network classifier the EEG segment (or the Fourier spectrum) of the subject which was not included into a training database, and to fix to which cluster of the trained classifier the EEG of the considered subject belongs.

Keywords: data mining, neural network classifiers, electroencephalograms (EEG), state of consciousness

Modeling of mortality of the Czech and Spanish populations

Petra Dotlacilova, Tomas Fiala, Jitka Langhamrova

Department of Demography, University of economics, Czech Republic

Mortality is recently very important topic, not only for demographers. It is due to its connection with a discussed topic of population aging. Based on the analysis of population mortality in European countries, it is clear that there is a gradual improvement in mortality. This development is due to the increasing level of medical care, and also the growing interest of people in a healthy lifestyle. This leads to a gradual lengthening of life expectancy at exact age x and thus to improve mortality rates. This development means that more and more people live to the highest ages. Previously, there was no need to pay close attention to the oldest-old mortality, because the highest ages lingered only a small number of people. Recently, the situation is changed. More and more people live to the highest ages. Still, however, that the oldest-old mortality is different from the mortality of younger persons. For this reason, it is necessary to model mortality of oldest persons. Previously, was the most used Gompertz-Makeham function. The development means that it is necessary to look for new models. Among the most popular are logistic models. In this paper we will analyze the mortality of Czech and Spanish population. The selected logistic models (Thatcher and Kannisto) will be applied on disposable data about mortality. The results will be compared with previously most common Gompertz-Makeham model. The aim is to determine which of these models the best describe mortality in both countries. The second aim is to determine if there are any changes during the reporting period. In the last part we will compare mortality trend in both countries.

Keywords: mortality, Gompertz-Makeham function, Kannistö and Thatcher model

The square-well model within the mean spherical approximation as a reference system in variational calculations for liquid metals

N.E. Dubinin, V.V. Filippov, A.A. Yuryev, N.A. Vatolin

Institute of Metallurgy of the Ural Branch of Russian Academy of Sciences, Russia

The variational method of the thermodynamic perturbation theory is widely used in structure and thermodynamics calculations for liquid and amorphous metals. Recently, the square-well model (SW) was suggested as a reference system in the variational calculations for simple liquid metals for the first time [1]. It was found that the SW-variational procedure gives a lower upper bound of the Helmholtz free energy than the hard-sphere-variational procedure. In [1] the SW system was taken within the random phase approximation (RPA). It is a shortcoming of the work [1] since the SW-RPA solution leads to unphysical behavior of the structure factor at some sets of the SW parameters. Progress in this area can be achieved by using more accurate theories of liquids, such as the mean spherical approximation (MSA).

Here, we use the SW-MSA reference system for which the semi-analytical approach developed in [2, 3] is applied. The needed expression for the entropy is obtained for this case.

Thermodynamics and structure of pure liquid alkali metals are studied. It is shown that the Helmholtz free energy (HFE) as a function of SW parameters has not a global minimum. Results obtained in local HFE minimums for each metal under consideration agree better with experiment than ones obtained earlier [1] with the SW-RPA reference system.

This work is supported by the Ural Branch of the Russian Academy of Sciences (grant RCP-13-P14), Russian Foundation for Basic Research (grant 11-03-01029-a), and Russian Program of Scientific Schools (grant 1278.2012.3).

References

1. N.E. Dubinin, A.A. Yuryev, N.A. Vatolin, J. Non-Equilibrium Thermodynamics, 35 (2010), p. 289.
2. N.E. Dubinin, V.V. Filippov, N.A. Vatolin, J. Non-Crystal. Solids, 353 (2007), p. 1798.
3. N.E. Dubinin, V.V. Filippov, O.G. Malkhanova, N.A. Vatolin, Cent. Eur. J. Phys., 7 (2009), p. 584.

Healthy Life expectancy and determinants of self-perceived health

In the three Baltic countries: commonalities and differences

Natalja Dubkova

University of Latvia, Latvia

Since 1991, when Estonia, Latvia and Lithuania regained independence from the Soviet Union, significant changes were experienced. Social and economic reforms of the 1990s have had a dramatic impact on almost all aspects of life – labour market, pension system and health-care system. With the independence renewal income inequality has increased substantially, but fast transition to market economy had a negative impact on population well-being and caused life expectancy decrease in many countries of post-communist space.

By present situation have improved significantly. Estonia became a leader among the Baltic countries in terms of life expectancy followed by Latvia and Lithuania. Population of Baltic countries like of other EU members experiences a systematic increase in life expectancy as well as in healthy life expectancy.

Analysis aims to give a summary of commonalities and differences in healthy life expectancy trends, health perception and factors behind the changes in the three Baltic countries - Estonia, Latvia and Lithuania.

Study involves analysis of differences in health perception by sex, age, education and main causes of differences in health perception.

In all the three Baltic countries males are likely to declare better health status than women, but significant differences in health perception in Latvia, Lithuania and especially Estonia had been observed.

Estonian citizens (especially women) more often than Latvian residents rate their health as good and very good. At the same time the share of reported long-term illness for both sexes in all age groups in Estonia over the Latvian and Lithuanian indicators is higher as well as the share of reported limitations in daily activities.

Self-perceived health of Latvian males and females evaluated with Chi Square Test of Independence showed, that with probability level of 99% there is a significant association between gender and self-perceived health status. Health condition, which is characterized by more or less serious limitations in daily activities, differs significantly among the age groups. If in younger ages about 9/10 of Latvian males and females reports on absence of any activity limitations, than in after retirement ages they are only 20-30%. The incidence of prolonged / chronic diseases grows with increasing age, peaking at around 70 percent of men and women aged 75 and over.

The last available indicator on healthy life years in the three Baltic countries was below the average of EU 27 by 3.9-8.2 years for males and 0.2-5.9 for females.

Proportion of healthy life years in the total life expectancy also is very significant indicator especially for retirement ages. In Latvia this indicator for 65+ aged population is 2-9 percent points lower than in Lithuania and Estonia.

Keywords: healthy life expectancy, Baltic countries, self-perceived health

Robust interpolation of periodically correlated random processes

Iryna Dubovetska, Mykhailo Moklyachuk

Department of Probability Theory, Statistics, and Actuarial Mathematics,
Kyiv National Taras Shevchenko University, Ukraine

We consider the problem of optimal linear estimation of the functional

$A_N \zeta = \int_0^{(N+1)T} a(t) \zeta(t) dt$ which depends on the unknown values of a periodically correlated stochastic process $\zeta(t)$, $t \in \mathbb{R}$, $[1]$, from observations of the process $\zeta(t)$ for $t \in \mathbb{R} \setminus [0, (N+1)T]$. After transformations we can write the functional in the form $A_N \zeta = \sum_{j=0}^N \vec{a}_j^T \vec{\zeta}_j$, where $\vec{\zeta}_j = (\zeta_{kj}, k=1, 2, \dots)$,

$\zeta_{kj} = \langle \zeta_j, \theta_k \rangle = T^{-1/2} \int_0^T \zeta(v + jT) e^{-2\pi i \{(-1)^k [\frac{k}{2}]\} v/T} dv$, is a vector sequence that corresponds to the stationary sequence $\{\zeta_j = \zeta(u + jT), u \in [0, T], j \in \mathbb{A}\}$, the coefficients $\vec{a}_j = (a_{kj}, k \geq 1) = (a_{1j}, a_{3j}, a_{2j}, \dots, a_{2k+1,j}, a_{2k,j}, \dots)^T$, $a_{kj} = \langle a_j, \theta_k \rangle$.

With the help of the Hilbert space projection method developed by A.N.Kolmogorov [2] we found the spectral characteristic and the mean square error of the optimal linear estimate of the functional $A_N \zeta$ under the condition that the spectral density $f(\lambda)$ of stationary sequence $\{\zeta_j, j \in \mathbb{A}\}$ is known.

In the case spectral uncertainty, where the spectral density $f(\lambda)$ is not known exactly, but, instead, a class of admissible spectral densities is given formulas that determine the least favorable spectral densities and the minimax (robust) spectral characteristics [3] of optimal estimates of the functional are proposed. It is shown that spectral densities of vector valued autoregressive stochastic sequences are least favorable for the optimal linear estimation of $A_N \zeta$ in certain rather general classes of admissible spectral densities.

References

1. E.G. Gladyshev. Periodically and almost periodically random processes with continuous time parameter, *Theory Probab. Appl.* 8, 173-177, 1963.
2. A.N. Kolmogorov. *Selected works of A.N. Kolmogorov. Vol. II: Probability theory and mathematical statistics*, Ed. by A.N. Shiriyayev. Mathematics and Its Applications. Soviet Series. 26. Dordrecht etc.: Kluwer Academic Publishers, 1992.
3. M.P. Moklyachuk, A.Yu. Masyutka, *Minimax-robust estimation technique. For stationary stochastic processes*, LAP LAMBERT Academic Publishing, 2012.

What drives the shape? – Cause of Death Contributions to Survival Shapes

Marcus Ebeling¹, Marie Pier Bergeron Boucher²

¹European Doctoral School of Demography and Max Planck Research Group Modeling the Evolution of Aging, Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany, ²European Doctoral School of Demography

The dramatic rise of life expectancy due to mortality declines across nearly all ages is one of the most remarkable demographic developments of the last century (Oeppen and Vaupel, 2002; Riley, 2001). The declining burden of infectious diseases and the decline of cardiovascular mortality are two salient trends underpinning the rise of life expectancy in the developed world (cf. Caselli et al., 2006; Omran, 1971). Overall, these trends indicate how shifting distributions of cause of death have significantly shaped survival and mortality (cf. Pollard, 1988, among others).

From a theoretical point of view, rising life expectancy means improved survival, as indicated by changing survival probabilities. These changes are reflected in the rectangularization of the survival curve - a process driven by an increasing number of people surviving to higher ages along with a decreasing variability of the age at death (Canudas-Romo, 2008; Cheung et al., 2005; Fries, 1980; Kannisto, 2000; Nusselder and Mackenbach, 1996; Vaupel and Zhang, 2008; Wilmoth and Horiuchi, 1999; Yashin et al., 2002).

While we know that shifting burdens of disease have contributed to changing survival probabilities, we have a limited understanding of how different causes of death contribute to the changing shape of the survival curve over time (cf. Myers and Manton, 1984; Nusselder and Mackenbach, 1997). In the study, we ask how and to what extent specific causes of death shape the survival curve using a new approach.

Based on cause-deleted survival curves, we measure rectangularization with the Maximized Inner Rectangle Approach (MIRA). Cause-deleted survival curves are extracted from cause elimination life tables (Beltran Sanchez et al., 2008). MIRA is a novel approach in survival shape measurement. It is based on a distinction between rectangularization of life expectancy (inner) and rectangularization towards the maximum living potential (outer). The measure decomposes maximum living potential into different areas, which enable a partial life expectancy examination. Determined by age a that maximizes $A(x) = x \cdot S(x)$, the application of MIRA is based on indexes measuring the horizontal, rectangular and longevity dimensions of a survival curve. Figure 1 shows MIRA and its application to a survival curve. The elimination of causes of death provides a measurement of how survival and the distribution of deaths are assigned to a particular cause of death (Beltran- Sanchez et al., 2008). MIRA allows an comprehensive study of the general effect as well as age patterns influencing the survival curve.

Applying MIRA to French and Canadian data, this study sheds new perspective on the impact of different causes of death on the shape of the survival curve across time.

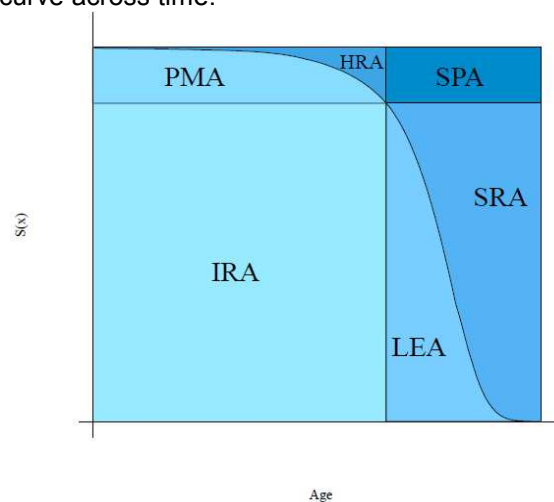


Figure 1: Scheme of the Maximized Inner Rectangle Approach (PMA-Premature Mortality Area, IRA-Maximized Inner Rectangle Area, LEA-Longevity Extension Area, HRA-Horizontal Rectangle Area, SPA-Shift Potential Area, SRA-Senescence Rectangle Area)

Keyword: MIRA, causes of death, rectangularization, survival curve, mortality

References

Beltran-Sanchez, H., S. H. Preston, and V. Canudas-Romo (2008). An integrated approach to causeof- death analysis: cause-deleted life tables

- and decompositions of life expectancy. *Demographic Research* 19(35), 1323–1350.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19(30), 1179–1204.
- Caselli, G., J. Vallin, and G. Wunsch (Eds.) (2006). *Demography - Analysis and Synthesis - A Treatise in Population Studies*, Volume 2, Chapter 57 - Health Transition: Trends and Prospects. Elsevier.
- Cheung, S., J.-M. Robine, E. Tu, and G. Caselli (2005). Three dimensions of the survival curve: horizontalization, verticalization, and longevity extension. *Demography* 42, 243–258.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3), 130–135.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6), –.
- Myers, G. C. and K. G. Manton (1984). Compression of Mortality: Myth or Reality? *The Gerontologist* 24(4), 346–353.
- Nusselder, W. J. and J. P. Mackenbach (1996). Rectangularization of the survival curve in the Netherlands, 1950–1992. *The Gerontologist* 36(6), 773–782.
- Nusselder, W. J. and J. P. Mackenbach (1997). Rectangularization of the survival curve in the Netherlands: An analysis of underlying causes of death. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 52B(3), S145–S154.
- Oeppen, J. and J. W. Vaupel (2002). Broken limits to life expectancy. *Science* 296(5570), 1029–1031.
- Omran, A. R. (1971). The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *The Milbank Memorial Fund Quarterly* 49(4), 509–538.
- Pollard, J. H. (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography* 25(2), pp. 265–276.
- Riley, J. C. (2001). *Rising Life Expectancy*. Cambridge University Press.
- Statistics Canada (2011). *Canadian vital statistics: death registration 1979–2008*.
- Vallin, J. and F. Meslé (2013). INED's database 1925–1999 (La base sur les causes de décès en France depuis 1925). online available. www.ined.fr.
- Vaupel, J. W. and Z. Zhang (2008). The Threshold between Comparison and Expansion of Mortality. paper distributed at the PAA 2008, 1–13.
- Wilmoth, J. and S. Horiuchi (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography* 36, 475–495.
- Yashin, A. I., A. S. Begun, S. I. Boiko, S. V. Ukraintseva, and J. Oeppen (2002). New age patterns of survival improvement in Sweden: do they characterize changes in individual aging? *Mechanisms of Ageing and Development* 123(6), 637 – 647.

Bias correction of Southwood & Jepson's graphical method of stage frequency data analysis by simulating *Eurygaster integriceps* data

V. Ebrahimi¹, S. Iranipour¹ and R. Karimzadeh¹

¹Department of plant protection, Faculty of agriculture, University of Tabriz, Tabriz, 51666-14888 Iran

Surveying number entering a particular stage of a stage-structured population (e.g. arthropods) using field sampling data has been subject of many studies in last 60 years. The Southwood and Jepson's graphical method (1962) is a simple and still applicable method in realistic situations in which the daily survival rate varies from stage to stage, or when a full set of data is not available for the entire generation. This method needs one beware of stage duration and the majority of mortality occurs at the end of a stage. Violation of the recent assumption leads to bias in estimations of this method. Some studies have been focused on improvement of the method assuming more realistic patterns of mortality distribution (e.g. Sawyer & Haynes, 1984). Assuming a constant rate of mortality and a separate estimate of recruitment to the first stage, Richards *et al.* (1961) proposed a similar method differing in pattern of mortality. In our study also similar assumption for mortality pattern was adopted. In absence of information about number entering the first stage, providing that the last stage reaches to zero (or alternatively, number of the last stage rather than the first stage be known), we could obtain more logical estimates with minor biases. Microsoft Excel was used to simulate stage frequency of six insect populations containing individuals passing through stages. The data simulation was based on information available by Iranipour *et al.* (2011) obtained on sunn pest, *Eurygaster integriceps* Puton (Hem., Scutelleridae). Different stage mortalities (10, 25, 50, 75, 90 and 99%) with constant daily rates were processed in the simulations. A laboratory reared cohort of Colorado potato beetle *Leptinotarsa decemlineata* Say (Col., Chrysomelidae) also was satisfactorily used to evaluate the method with more realistic patterns of mortality rather than a constant rate. In absence of any information about number entering last or first instars (for example when last stage begins diapause, migrate or fly out prior to death and do not remain more for sampling up to end of a generation), Sawyer and Haynes method is advantageous. When data are present for a whole generation terminating by dying the last individual, our method will be preferable.

Keywords: Stage-frequency data, Survival rate, Simulation, Mortality rate.

On Option Pricing in Illiquid Markets with Jumps

Youssef El-Khatib¹, Abdunnasser Hatemi-J²

¹Department of Mathematical Sciences, UAE University, ²Department of
of Economics and Finance, UAE University, United Arab Emirates

One of the shortcomings of the Black and Scholes model on option pricing is the assumption that trading of the underlying asset does not affect the price of that asset. This assumption can be fulfilled only in perfectly liquid markets. Since most markets are illiquid, this assumption might be too restrictive. Thus, taking into account the price impact in option pricing is an important issue. This issue has been dealt with, to some extent, for illiquid markets by assuming a continuous process, mainly based on the Brownian motion. However, the recent financial crisis and its effects on the global stock markets have propagated the urgent need for more realistic models where the stochastic process describing the price trajectories involves random jumps. Nonetheless, works related to markets with jumps are scant compared to the continuous ones. In addition, these previous studies do not deal with illiquid markets. The contribution of this paper is to tackle the pricing problem for options in illiquid markets with jumps as well as the hedging strategy within this context, which is the first of its kind to the best knowledge.

Keywords: Options Pricing, Illiquid Markets, Jump Diffusion, Incomplete Markets.

A Double HMM Approach to Altman Z-scores and Credit Ratings

Robert J. Elliott^{1,2}, Tak Kuen Siu³, Eric S. Fung⁴

¹School of Mathematical Sciences, University of Adelaide, Australia,

²Haskayne School of Business, University of Calgary, Canada,

³Department of Applied Finance and Actuarial Studies, Faculty of
Business and Economics, Macquarie University, Australia, ⁴Department
of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong
Kong

Credit ratings and accounting-based Altman Z-scores are two important sources of information for assessing the creditworthiness of firms. In this paper we build a model based on a double Hidden Markov Model, (DHMM), to extract information about the "true" credit qualities of firms from both the Z-scores evaluated from the accounting ratios of the firms and their posted credit ratings. The evolution of the "true" credit quality over time is estimated from observed data using filtering methods and the EM algorithm.

Recursive updates of optimal estimates are provided via filtering so that the model is "self-tuning", or "self-calibrating". We illustrate the practical implementation of the proposed model using actual accounting ratios data of firms from different regions and their posted credit ratings data.

Keywords: Accounting-based credit scoring systems; Altman Z-scores; Credit Ratings; Double Hidden Markov Models; Reference Probability; Filters; EM algorithm

Variations of PageRank with application to linguistic data

Christopher Engström¹, Sergei Silvestrov¹, and Thierry Hamon²

¹Division of Applied mathematics, Education, culture and communication, Mälardalen University, Sweden

²LIM&BIO (EA3969), Université Paris 13, Sorbonne Paris, France

In this paper we will give some alternative formulations of PageRank originally used by Google to rank homepages in a search query. By looking at PageRank from a probabilistic perspective we will see how to fix some of the limitations in the original definition of PageRank while still retaining the main strengths of the algorithm. Namely we will see that the alternative version can still be computed using the same methods (power iteration) but without the dependence on graph size in the original definition. This makes it possible to compare rank not only within a system, but between different systems as well. A short comparison between this alternative definition and the original definition is given accompanied with proofs of how to get from one definition to the other using some simple calculations. Last we will take a short look at an application of PageRank to linguistic data for clustering of terms or identification of "important" terms in a corpus of biomedical texts. We show how the sub-stochastic system matrix can be created for this kind of data as well as highlighting some of the limitations of the method and future work.

Keywords: PageRank, Markov chains, text mining.

DECREASE IN RISK OF ERRONEOUS CLASSIFICATION OF THE MULTIVARIATE STATISTICAL DATA

Farhadzadeh E.M., Farzaliev Y.Z., Muradaliev A.Z.

Azerbaijan Scientific-Research and Design-Prospecting Institute of Energetic, Russia

Estimation of parameters of individual reliability of the equipment of power supply systems provides classification of final population of multivariate statistical data of operation, tests and restoration of

deterioration on the set versions of attributes. Versions of attributes reflect features of a design, a condition of operation, feature of occurrence of refusals and carrying out of repairs of the equipment. The expediency of classification on each of set versions of attributes is established by comparison of statistical functions of distribution of final population of statistical data and statistical function of distribution of sample n random variables from this set $F_{V,i}^*(X)$. In a basis of comparison $F_{\Sigma}^*(X)$ and $F_V^*(X)$ there is a statistical modeling n pseudo-random numbers ξ , With uniform distribution in an interval $[0,1]$. An indispensable condition thus is consistency of statistical function of distribution $F_V^*(\xi)$ to the uniform law of distribution $F_{\Sigma}(\xi)$. It is obvious, that from the uniform law of change of random numbers ξ At all consistency does not follow the uniform law of statistical function of distribution $F_V^*(\xi)$ with the set significance value α .

The interval nonparametric criterion of the control of conformity выборок from n pseudo-random numbers is offered to the uniform law in an interval $[0,1]$. In a basis of criterion there is a distinction of distributions of positive and negative values of the greatest divergence of distributions $F_{\Sigma}(\xi)$ and $F_V^*(\xi)$. Transition from statistics D_n to statistics St_n allows not only to simplify algorithm of calculation of the greatest divergence $F_{\Sigma}(\xi)$ and $F_V^*(\xi)$, but also to estimate an opportunity of use of statistics St_n at an estimation of the greatest divergence of statistical function of distribution $F_{\Sigma}^*(X)$ and $F_V^*(X)$, to estimate risk of the erroneous decision.

Increase of accuracy of the control of conformity of distribution $F_V^*(\xi)$ to the uniform law reached by practical realization of recommended algorithm of the decision-making considering not only a mistake of the first sort, but also the second sort.

Kernel type estimator of a bivariate average growth function

István Fazekas¹, Zsolt Karácsony², Renáta Vas¹

¹Faculty of Informatics, University of Debrecen, ²Faculty of Mechanical Engineering and Informatics, University of Miskolc, Hungary

In the paper, the usual regression problem is considered and a bivariate function with general error is observed. The Gasser-Müller method is used to obtain the estimator of the unknown function. Under general and realistic conditions on the covariance structure of the error random field

an upper bound is obtained for the mean squared error. Simulation results are also presented.

Keywords: Average growth function, kernel, Gasser-Müller estimator, Mean squared error, Bandwidth, Taylor expansion.

References:

- (1) K. Benhenni and M. Rachdi: *Nonparametric estimation of average growth curve with general nonstationary error process*. (2007) (Communications in Statistics - Theory and Methods, 36, 1173-1186.)
- (2) T. Gasser and M.G. Müller: *Estimating regression functions and their derivatives by the kernel method*. (1984) (Scandinavian Journal of Statistics, Theory and Applications, 11, 171-185.)

The publication was supported by the TAMOP-4.2.2.C-11/1/KONV-2012-0001 and TAMOP-4.2.2/B-10/1-2010-0024 projects. The projects have been supported by the European Union, co-financed by the European Social Fund.

Scale-free property for degrees and weights in a preferential attachment random graph model

István Fazekas, Bettina Porvázsnyik

University of Debrecen, Faculty of Informatics

During the last 15 years the behaviour of many type of real-world networks was investigated. Such networks are the WWW, the Internet, social and biological networks (see [1] for an overview). The main common characteristic of such networks is their scale-free property, in other word the power law degree distribution. To describe the phenomenon, in [2] the preferential attachment model was introduced. In [3] a model based on the interaction of three vertices was introduced. Then the power law degree distribution in that model was obtained in [4]. In this paper, a preferential attachment random graph model based on the interaction of four vertices is studied.

We extend the model and the results of [3] and [4] to four interactions. Scale free properties for weights and degrees are obtained.

Keywords: Random graph, preferential attachment, scale free, power law, submartingale, Doob-Meyer decomposition.

[1] Durrett, R. Random graph dynamics. Cambridge University Press, Cambridge UK, (2007).

[2] Barabasi, A.L., Albert, R. Emergence of scaling in random networks. Science 286 (1999), 509-512.

[3] Backhausz, A., Mori, T. F. A random graph model based on 3 interactions. Annales Univ. Sci. Budapest., Sect. Comp. 36 (2012), 41-52.

[4] Backhausz, A., Mori, T. F. Weights and degrees in a random graph model based on 3-interactions. arXiv:1206.0633v1 [math.PR] 4 Jun 2012.

The publication was supported by the TAMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

The distribution of the average lifetime of the Marshall–Olkin law

Lexuri Fernández¹, Matthias Scherer²

¹**University of the Basque Country UPV/EHU, Ekonomia Analisaren**
Oinarriak II Saila, Spain, ²Lehrstuhl für Finanzmathematik (M13),
Technische Universität München, Germany

The probability distribution of $S_d := X_1 + \dots + X_d$, where (X_1, \dots, X_d) is distributed according to the Marshall–Olkin law, is investigated. The distribution of $S_d := X_1 + \dots + X_d$ has been considerably treated in the literature. For mathematical tractability, the individual random variables X_k are often considered to be independent, see, e.g., [2], an hypothesis that is hardly never met in real-world applications. In our study we assume (X_1, \dots, X_d) to be distributed according to the Marshall–Olkin law, a popular assumption for dependent lifetimes in insurance and credit risk modeling. With this interpretation in mind, S_d/d denotes the average lifetime of dependent exponential random variables.

[11] introduce a d -dimensional exponential distribution by lifting the univariate lack of memory property $P(X > x + y | X > y) = P(X > x)$, for all $x, y > 0$, to higher dimensions. More precisely, it is shown that the only distribution with support $[0, \infty)^d$ that satisfies the multidimensional lack of memory property

$P(X_1 > x_1 + y, \dots, X_d > x_d + y | X_1 > y, \dots, X_d > y) = P(X_1 > x_1, \dots, X_d > x_d)$, (1)

where $x_1, \dots, x_d, y > 0$, is characterized by the survival function of the Marshall–Olkin distribution,

$$\bar{F}(x_1, \dots, x_d) = P(X_1 > x_1, \dots, X_d > x_d) = \exp \left(- \sum_{i \in I} \lambda_i \max_{j \in I} \{x_j\} \right)$$

such that $\lambda_1 \geq 0$, $\emptyset \neq I \subset \{0, 1, \dots, d\}$, and $\sum_{k \in I} \lambda_k > 0$, $k = 1, \dots, d$.

Interpreting X_k as lifetime of component k , $\{\lambda_1 : I \subset \{1, \dots, d\}\}$ represent

intensities of “shocks” influencing the lifetime of all components in I . This can be seen from the canonical construction of the Marshall–Olkin distribution that is motivated as a fatal-shock model, see [10, 6].

We derive $P(S_2 > x)$ explicitly in the general case. The involved computations can be extended to higher dimensions, which, however, becomes cumbersome due to the large number of involved parameters. One property of the Marshall–Olkin law is the large number of parameters, namely $2^d - 1$ in dimension d , rendering the Marshall–Olkin challenging to work with as d increases. To account for this, we focus on the exchangeable subfamily, which only has d parameters in dimension

d . The distribution of $P(S_d > x)$ is found for $d \in \{2, 3, 4\}$.

We guide the interested reader to strategies how extensions to higher dimensions might be achieved. Moreover, we study the asymptotic distribution of S_d/d , when $d \rightarrow \infty$, in the extendible Marshall–Olkin subfamily. Via a detour to Lévy-frailty copulas, see [9], this is related to certain exponential functionals of Lévy subordinators, see [7, 3, 13, 8]. This result might serve as a convenient approximation in high-dimensional cases.

Related studies on the probability of a sum of dependent risks can be found in the literature related to insurance, in the case of a natural catastrophe ([5]), and risk-management, see, e.g., ([1], [12], [14], [4]).

References

- [1] P. Arbenz, P. Embrechts, and G. Puccetti. The aep algorithm or the fast computation of the distribution of the sum of dependent random variables. *Bernoulli* Bethesda, 17(2):562, 2011.
- [2] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- [3] J. Bertoin, M. Yor, et al. Exponential functionals of Lévy processes. *Probab. Surv.*, 2:191–212, 2005.
- [4] H. Cossette and É. Marceau. The discrete-time risk model with correlated classes of business. *Insurance: Mathematics and Economics*, 26(2):133–149, 2000.

- [5] M. Denuit, C. Genest, and É. Marceau. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85–104, 1999.
- [6] F. Downton. Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society. Series B, (Methodological)*, pages 408–417, 1970.
- [7] H.K. Gjessing and J. Paulsen. Present value distributions with applications to ruin theory and stochastic equations. *Stochastic processes and their applications*, 71(1):123–144, 1997.
- [8] A. Kuznetsov and Pardo J.C. Fluctuations of stable processes and exponential functionals of hypergeometric lévy processes. *Acta Applicandae Mathematicae*, pages 1–27, 2010.
- [9] J.F. Mai and M. Scherer. Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585, 2009.
- [10] Albert W. Marshall and Ingram Olkin. A generalized bivariate exponential distribution. *Journal of Applied Probability*, pages 291–302, 1967.
- [11] A.W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44, 1967.
- [12] G. Puccetti and L. Rüschendorf. Sharp bounds for sums of dependent risks. preprint, 2012.
- [13] V. Rivero et al. Tail asymptotics for exponential functionals of lévy processes: the convolution equivalent case. arXiv preprint arXiv:0905.2401, 2009.
- [14] MVWuthrich. Asymptotic value-at-risk estimates for sums of dependent random variables. *Astin Bulletin*, 33(1):75–92, 2003.

A simulated approach of Bayesian variability of adjusted premium for extremes values

Kmar FERSI¹, Kamal BOUKHETALA², Samir BEN AMMOU³

¹Institut Supérieur de Gestion, Sousse & Faculté des Sciences, Computational Mathematics Laboratory, Tunisia, ²Faculté de Mathématiques Bp. 32, USTHB, Algeria, ³Faculté des Sciences, Computational Mathematics Laboratory, Tunisia

Basing on the Wang's premium principle, Necir A. and Boukhetala K. (2004) have proposed an estimator of risk-adjusted premium for extremes claims amounts which represented in the heavy-tailed distributions. In this paper, we propose a simulated approach for a Bayesian variability of the adjusted premium using GPD fit, for taking into account the ambiguity effect of these extremes values (Keller and al. (2007)). Moreover, for given a high threshold, we calculate the adjusted premium with a Bayesian form respect to parameter of distortion which represents the ambiguity effect (ambiguity aversion) and we propose a

stochastic optimization program for reducing the variability of actuarial premiums. For a reason of difficulty to obtain an exact solution of the problem, an optimization hybrid genetic algorithm is used. An application to automobile insurance is developed.

Keywords: Premium calculation, Wang's premium, ambiguity aversion, Extreme value, GPD function, hybrid genetic algorithm, optimization.

References

1. Cebrian, A., Denuit M. and Lambert P. (2004). Generalized Pareto Fit to the Society of Actuaries Large Claims Database, North American Actuarial Journal, 7, 18-36.
2. Embrechts P., Kluppelberg C. et Mikosch T. (1997). Modelling Extremal Events for Insurance and Finance. Springer, Berlin.
3. Johansson J. (2003). Estimating the Mean of Heavy-Tailed distributions. Extremes, 6, 91-109.
4. Keller, R., Sarin R., and Souderpandian J. (2007): An examination of ambiguity aversion. Judgment and Decision Making, 5, 390-397.
5. Necir, A. and Boukhetala K. (2004). Estimating the Risk-Adjusted Premium for The Largest Claims Reinsurance Covers. COMPSTAT'2004. Proceeding in Computational Statistics. ISBN 3-7908-1554-3, Physica- Verlag, Heidelberg, New York.
6. Pickands, J. (1975). Statistical inference using extreme order statistics. Ann. Statist., 3, 119-131.

Modeling of Mortality in Older Ages by a Modified Gompertz-Makeham Function

Tomas Fiala, Petra Dotlacilova, Jitka Langhamrova

Department of Demography, Faculty of Informatics and Statistics,
University of Economics, Czech Republic

One of the functions very often used for modeling of mortality at higher ages is the Gompertz-Makeham function

$$\mu(x) = a + b \cdot c^x,$$

where x denotes the age, $\mu(x)$ is the force of mortality, a , b , c are parameters of the function.

Computations for many countries show that for higher old age (80 years and more) the values of Gompertz-Makeham function are systematically higher than the empirical values of age-specific death rates.

Koschin (1998) proposed to modify the Gompertz-Makeham function for higher old age in the following way

$$\mu(x) = a + b \cdot c^x, \quad x \leq 83; \quad \mu(x) = a + b \cdot c^{83 + \frac{1}{d} \cdot \ln[d \cdot (x-83) + 1]}, \quad x > 83,$$

where $d > 0$ is additional parameter.

The paper will bring computations of modeling mortality of persons of 60 years of age and older by the „classical“ Gompertz-Makeham function and by its modified form. The estimate of unknown parameters of the function will be based on the weighted least square method. Computations will be carried out in Excel by the procedure *Solver* for solving nonlinear equations. Results for various countries will be presented.

Keywords: mortality, force of mortality, age-specific mortality rates, Gompertz-Makeham function

Koschin, F., Fiala, T., Langhamrová, J., Roubíček, V. *Mortality in the Czech Republic in the Nineties*. 1. vyd. Praha : VŠE, 1998. 70 pp. ISBN 80-7079-586-7.

Some Results on Generalized Mixtures of Weibull Distributions with Different Shape Parameters*

Manuel Franco¹, Debasis Kundu², Narayanaswamy Balakrishnan³, Juana-Maria Vivo⁴

¹Department of Statistics and Operations Research, University of Murcia, Spain, ²Department of Mathematics & Statistics, Indian Institute of Technology Kanpur, India, ³Department of Mathematics and Statistics, McMaster University Hamilton, Canada, ⁴University of Murcia, Regional Campus of International Excellence "Campus Mare Nostrum", Spain

Weibull distributions and their mixtures play a great role in the reliability theory to model lifetime and failure time data, since they can incorporate wide varieties of failure rate functions. These mixtures forms have been generalized by allowing negative mixing weights, which arise under the formation of some structures of systems, and provide more versatile distributions for modelling dependent lifetimes from heterogeneous populations, see, for example, Jiang et al. [8].

Moreover, the generalized mixtures of Weibull distributions can be considered as extensions of the generalized mixtures of exponentials, which have been characterized in terms of the mixing weights and the parameters of its exponential components, and related results have been obtained by several authors, among others, see Baggs and Nagaraja [1], Bartholomew [2], Franco and Vivo [3-5] and Steutel [9]. In this setting, Franco and Vivo [6] and Franco et al. [7] have expanded the study of generalized mixtures of exponential components to the case of Weibull components with a common shape parameter.

In this work, we propose to study generalized mixtures of Weibull distributions with different shape parameters, by establishing its characterization in terms of the mixing weights and the parameters of its Weibull components to be a valid probability model.

***This work was partially supported by Fundacion Seneca of the Regional Government of Murcia (Spain) under Grant 11886/PHCS/09.**

Keywords: Weibull distribution, Mixture model, Generalized mixture

References

- [1] G.E.Baggs and H.N. Nagaraja. Reliability properties of order statistics from bivariate exponential distributions, Comm. Statist. Stochastic Models 12, 611–631, 1996.
- [2] D.J. Bartholomew. Sufficient conditions for a mixture of exponentials to be a probability density function, Ann. Math. Statist., 40, 2183–2188, 1969.

Multiple kernel testing procedures with non-asymptotic bootstrap approaches for two-sample problems

**Magalie FROMONT, Béatrice LAURENT, Matthieu LERASLE,
Patricia REYNAUD-BOURET**

CREST (ENSAI) - IRMAR (Université Rennes 2) ENSAI, FRANCE

Considering either two independent Poisson processes or two independent i.i.d. samples, we address the question of testing equality of their respective distributions.

We first propose testing procedures whose test statistics are U-statistics based on a single kernel function, which may be chosen as a projection, approximation or reproducing kernel. The corresponding critical values are constructed from specific non-asymptotic bootstrap approaches, as a result the obtained tests are exactly (and not only asymptotically) of level α .

We then introduce aggregated or multiple kernel testing procedures, which enable us to overcome the difficulty of choosing the kernel and/or the parameters of the kernel, and which allow us to import ideas coming from model selection, thresholding and/or approximation kernels adaptive estimation. These multiple kernel tests are proved to be exactly of level α , and to satisfy non-asymptotic oracle type inequalities leading to minimax adaptivity properties in some cases.

Keywords: Two-sample problem, multiple kernel methods, Poisson process, density model, wild bootstrap, permutation test, adaptive tests, aggregation methods

Continuous Logarithmic random variables and Exponential AR (1) process

Veeresh Gadag

Division of Community Health and Humanities, Faculty of Medicine,
Memorial University St. John's, Canada

We introduce a new class of probability distribution called continuous logarithmic distribution denoted by $L_C(\cdot; \beta, \theta)$ on $(0, \infty)$; where $0 < \beta (\neq 1) < \infty$ and $0 < \theta < \infty$.

We denote the class of distributions, whose *pdfs* are given by (1) as $L_C(\cdot; \beta, \theta)$. We show that its Laplace Transform is given by $\phi_X(s) = \{\log [(\beta\theta + s)/(\theta + s)]\}/(\log \beta)$. We note that these distributions can be used to fit data on continuous non negative random variables that are skewed with thick tails.

Using these continuous logarithmic *rvs*, we develop non-Gaussian autoregressive (AR) processes. We provide an interesting interpretation such an AR(1) process with reference to single server queue with 'batch arrivals'. We propose to obtain properties of this queue, such as, distribution of the busy service time and its expected value, expected number of customers receiving the service in the busy period among others. We also propose to develop several other different ways of building the exponential AR(1) processes generalizing the results of Lawrance and Lewis(1981), Sim (1990), Novkovic(1999) among others.

Optimal turbines layout in an offshore wind farm using evolutionary computation

D. Gallo-Marazuela, S. Salcedo-Sanz, A. Pastor-Sánchez, E. Alexandre-Cortizo, A. Portilla-Figueras, L. Prieto

Department of Signal Theory and Communications, Escuela Politécnica Superior, Universidad de Alcalá, Spain

The placement of the whole set of turbines of an offshore wind farm in northern Europe is tackled in this paper by means of evolutionary algorithms. We consider a new model where the shape of the wind farm is taken into account, an a new model of wake is applied using an existing simulator called *OpenWind*. The proposed evolutionary algorithm works by choosing the places where the turbines must be

installed (wind farm layout), in such a way that the energy produced by the wind farm is maximized. Different variants of the problem have also been taken into account, such as limiting the number of turbines to a fix number or limiting the number of turbine sites available, so a minimum distance between turbines is required to guarantee safety criteria. The results obtained have shown that the evolutionary approach is able to obtain excellent layouts, that maximize the power generation offered by the wind farm.

Construction of a happiness index using polytomous item response theory models in a survey

Diana Carolina Gamboa, Álvaro Mauricio Montenegro

Universidad Nacional de Colombia, Colombia

In this paper, we present an alternative to derive univariate indices by using polytomous item response theory models in data from surveys. Particularly, the Samejima's graded response model was used. A real data set from the Social European Survey database was fitted. In order to interpret the index a multiple correspondence analysis and a clustering was performed.

Keywords: Univariate index, polytomous item response theory, the European Social Survey, multiple correspondence analysis.

Nonparametric estimation of the intensity function of a trend renewal process

M.L. Gamiz, B.H. Lindqvist

Department of Statistics and O.R., University of Granada, Spain

Repairable systems are systems that are repaired to satisfactory performance after a failure. Traditionally, the literature on repairable systems treats the failure times using point process theory. The most common models for the failure process of a repairable system are nonhomogeneous Poisson processes (NHPP), related to what is called minimal repairs, and renewal processes (RP), related to perfect repairs or replacements. For many applications it is more reasonable, however, to model the repair action by something in between the two given extremes, leading to the need for more general models.

The trend-renewal-process (TRP) is a model of this kind, defined to be a time-transformed renewal process, where the time transformation is

given by a trend function similar to the intensity of a non-homogeneous Poisson process (NHPP).

In some sense the TRP is constructed as the "least common multiple" of the NHPP and the RP. Thus the TRP framework can be used to distinguish between the two "extreme" kinds of repair, minimal and perfect. Secondly, the TRP represents a possible trend in inter-failure times.

The present paper is concerned with the fitting of TRP models to failure data. Until now, this has been mostly done using parametric models for the trend function and the underlying renewal process.

However, in many applications there is no clear reason to choose a concrete parametric model for the conditional intensity function and in such a case a free-model method and then a data-driven focus of the problem seems to be the most adequate.

We consider a nonparametric approach and estimate the conditional intensity function by using kernel smoothing techniques. We develop an algorithm to estimate the intensity function by preserving its structure in terms of the trend function and the underlying renewal process.

Keywords: repairable systems; trend renewal process; kernel smoothing

On the Application of an Evolutionary Algorithm for the Optimization of Kitting Areas in Car Manufacturing Production Chains

C. A. Garcia-Santiago¹, A. Gonzalez-Gonzalez¹, I. Landa-Torres¹, J. Del Ser¹, S. Gil-Lopez¹, I. Diaz-Iriberry², F. Duran Limon, and S. Salcedo-Sanz³

¹TECNALIA. OPTIMA Unit, ²Nissan Motor Ibérica S.A., ³Universidad de Alcalá, Spain

Recently a flurry of research has been invested towards assessing and reducing the work saturation of operators in massive car production plants. Saturation not only incurs severe drawbacks in the effective productivity of the personnel at hand, but may also cause risky situations due to the reduced awareness of the operator with respect to the surrounding machinery. In order to prevent such risks, and to improve the operators' work conditions, a maximum saturation % is defined. For this reason, the companies in the sector are obliged to improve their processes as much as possible in order to ensure an optimum saturation level meaning, as a direct consequence, that their resources must be accurately allocated. Based on this rationale, the adoption of innovative approaches for the cost-less optimization of operational processes on-

site is deemed of utmost importance for the viability of the manufacturing company at hand.

Among the technical strategies utilized for avoiding this undesired saturation regime, the use of avant-garde optimization algorithms has lately gained momentum in the research community. Their use in the context of the optimization of production chains is not new, as many methods have been published in the literature dealing with production scheduling and lot storing, among many others. However, less attention has been paid to the optimization of the physical configuration of the production machinery, mainly due to the difficulty of dynamically moving steady machinery from one location to another, and the inherent economic costs derived therefrom.

However, modern car manufacturing chains require dynamic support systems in order to cope with the heterogeneity of parts and pieces of different product models. One of such machinery is the so-called Automated Guided Vehicle (AGV), which essentially consists of a moving trolley looping through a kitting area which serves as a buffer for the operator to pick and deploy the necessary parts to the assembly line. As shown in Figure 1, the kitting area is further composed by shelves of different capacity and size, which are set on specific reserved slots or PDLs (Point of Delivery at Line) containing different types of parts. Once the parts have been assigned and deployed in the shelves, the AGV is filled with items as a function of the model to be produced at a certain time. Interestingly for the scope of this work, the model series to be produced is reflected in the production forecast, which is usually fixed for a long term period. In this scenario, saturation comes from the fact that the operator in the kitting area must pick and deploy parts in the AGV before the trolley finishes the path, an intensive activity no matter if the production chain is producing at its maximum rate.

This being said, this research work gravitates on the use of modern evolutionary meta-heuristics as an efficient computational means to improve the saturation level in this practical scenario. A thorough mathematically modeling of the problem is presented, whose complexity motivates the use of derivative-free computational methods such as the herein proposed evolutionary meta-heuristics. Additional side algorithms and methods are also incorporated to the main evolutionary solver so as to account for operational constraints such as the number of sub-paths to be allowed for the AGV in its path through the kitting area, or the maximum number of parts in any of the installed shelves. The paper also discusses several numerical experiments aimed at verifying the saturation decrease rendered by the proposed algorithm with respect to strategies where the picking frequency of parts is adopted as a criterion for their deployment in the shelves.

**New approaches to study historical evolution of mortality
(with implications for forecasting)**

Gavrilov L.A., Gavrilova N.S

Center on Aging, NORC at the University of Chicago, USA

Increasing numbers of people surviving to advanced ages pose serious challenge to government pension systems and to the most industrialized societies. In the majority of developed countries fertility already reached very low levels and has little chances of radical changes, so mortality and mortality at advanced ages in particular is the main driving force behind the future population changes. Therefore, accurate estimates of mortality at advanced ages are essential to improving forecasts of mortality and the population size of the oldest old age group. In this talk we present some new approaches to mortality and population projections at older ages. We apply modified method of mortality shifting to population of Sweden and make mortality projections up to 2070. Specifically, we identify the best time interval for determining the rate of mortality decline, which is used in mortality extrapolation. In the case of Sweden, the best interval is 1980 through 2008 years for both men and women. For men, the rate of mortality decline was almost twice as high as the rate for female. Using two simple assumptions (log-linear decline of mortality over time and Gompertz law working up to very advanced ages), we made mortality projections for Swedish males and females for the next 60 years. According to these projections, life expectancy at age 25 will increase from 54.07 in 2005 to 62.71 in 2050 for men and from 58.20 to 63.50 for women. If this tendency of mortality decline continues then in 2059 life expectancy at age 25 for men may surpass that of women. These advances in life expectancy will not result in population growth and in the absence of migration the 'native' population of Sweden is expected to decline after 2036 (assuming unchanged birth rate).

Additional reading:

Gavrilova N.S., Gavrilov L.A. Ageing and Longevity: Mortality Laws and Mortality Forecasts for Ageing Populations [In Czech: Stárnutí a dlouhověkost: Zakony a prognózy umrtnosti pro stárnoucí populace]. *Demografie*, 2011, 53(2): 109-128.

Full text available at:

<http://longevity-science.org/pdf/Demografie-English-2011.pdf>

Quantitative Methods in Demography Workshop and Tutorial (QMDWT)

Mortality measurement and modelling at advanced ages

Gavrilova N.S., Gavrilov L.A.

Center on Aging, NORC at the University of Chicago, USA

The growing number of persons living beyond age 80 underscores the need for accurate measurement of mortality at advanced ages. Earlier studies suggested that the exponential growth of mortality with age (Gompertz law) is followed by a period of deceleration, with slower rates of mortality increase. This study challenges earlier conclusions with new data from the U.S. Social Security Administration's Death Master File and survival records for laboratory mice and rats. Analyses of human extinct birth cohorts (1890-1898) demonstrated that mortality deceleration is far less pronounced when it is measured over monthly (rather than yearly) age intervals, and in higher-quality data. Mortality deceleration is also challenged by recent rodent data. Simulations show that some estimates of mortality may produce spurious mortality deceleration, while the Sacher estimate turns out to be the most accurate estimate of hazard rate.

Additional reading:

Gavrilov L.A., Gavrilova N.S. Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File. *North American Actuarial Journal*, 2011, 15(3): 432-447.

Full text available at:

<http://longevity-science.org/pdf/Mortality-NAAJ-2011.pdf>

The Illumination Inverse Problem Applications to Electron Microscopy

Segolen Geffray¹, Nicolas Klutchnikoff², Myriam Vimond³

¹Université de Strasbourg et CNRS, Institut de Recherche Mathématique Avancée, ²CREST (Ensai et UEB), IRMA (UMR 7501 CNRS et

Université de Strasbourg), ³CREST (Ensai et UEB), IRMAR (UMR 6625 CNRS), France

We are interested in a multi-dimensional signal denoted by R that is observed with an illumination artefact inherent in the recording technology. This may be the case with electron microscopy for example. The illumination artefact is modelled by a function L which is "smooth" enough, in a sense to be precised in the mathematical developments, and which acts in a multiplicative way on the original signal R . We also assume the presence of an additional additive noise ε so that the

observed signal denoted by Y is actually linked to the original signal R by the equation: $Y(x) = R(x)L(x) + \varepsilon(x)$, for x describing the observation domain. In this framework, our aim consists of reconstructing R from the observation of Y . To this aim, we use a regression strategy. We project $\log Y$ on a suitable closed convex subspace of a Hilbert space so that we get an estimation of $\log L$ and deduce an estimation of R . The procedure quality is studied through the evaluation of the L2-risk of the estimator of R . Simulations are presented. At last, an application to real electron microscopy images is presented.

Keywords: Semi-parametric estimation, inverse problem, Retinex theory, color perception.

Reliability Evaluation of 2D Barcode OCR by using Monte-Carlo Simulation

Evgeny Gershikov, Samuel Kosolapov

Signal and Image Processing Laboratory, ORT Braude Academic
College of Engineering, Israel

Modern smartphones enable fast and easy recognition of a single two-dimensional (2D) Barcode presented to the smartphone' camera. In our current research, which deals with a camera-based Instant Feedback System (IFS), it is required to recognize multiple 2D barcodes presented to the camera. Additional IFS requirement is to evaluate the orientation of the 2D barcode relative to the camera. In order to make recognition of a plurality of barcodes, presented at the same time to the digital camera, feasible and reliable, special custom format of 2D barcode was created. Well-known OCR (Optical Character Recognition) algorithms were adapted to the multiple target recognition. In order to evaluate the feasibility of the selected IFS approach and selected 2D barcode design, MATLAB-based simulations were performed. The simulations covered the whole process starting from the acquisition of an IFS image and up to the recognition of the 2D barcodes. Considering a big number of the simulation parameters, the usage of the classical Monte-Carlo approach is necessary in order to evaluate the accuracy and reliability of the selected 2D barcode design and selected OCR algorithms. A number of practically interesting IFS configurations were analyzed and compared with respect to reliability. We conclude that the proposed IFS method is feasible at simulation level, but requires higher reliability OCR techniques for practical use. The proposed and similar techniques are also examined in real images.

Keywords: Image Processing, Instant Feedback System, OCR, 2D Barcode, Monte-Carlo Simulation

Relative Entropy Versus Entropy Difference in Goodness-of-Fit Tests. Application to Pareto fitting

Valérie Girardin¹

¹ Laboratoire de Mathématiques N. Oresme, UMR6139, Campus II, Université de Caen Basse Normandie, BP5186, 14032 Caen, France

Goodness-of-fit tests constitute a classical tool in deciding of the compatibility of data with some theoretical probability distribution. A large choice of such tests is available in the literature. Tests based on Shannon entropy or relative entropy (also called Kullback-Leibler information or divergence), that is entropy-based tests, will be especially considered and compared in this talk, in particular with the aim of testing the fitting of data with Pareto distributions.

Keywords: goodness-of-fit test, Kullback-Leibler divergence, Pareto distribution, relative entropy, Shannon entropy.

Reference

Valerie Girardin and Justine Lequesne Relative Entropy Versus Entropy Difference in Goodness-of-Fit Tests. Application to Pareto fitting. *Rapport technique du LMNO, UCBN, France* (2013).

Computation of Generalized Entropy Rates and Generalized Relative Entropy rates

Valerie Girardin¹, Loïck Lhote²

¹Laboratoire de Mathématiques N. Oresme, Université de Caen,

²Laboratoire GREYC, UMR CNRS 6072, Université de Caen, ENSICAEN, France

Recently, Ciuperca et al. (2011) proved that the generalized entropy rates of a wide class of random sequences are, under some conditions, zero or infinite except at a threshold where they are equal to the classical Shannon or Rényi rates (up to a multiplicative factor). This behavior, which limits their practical interest in applications, is due to the rescaling factor $1/n$ in the classical definition of the entropy rates. Based on a convenient rescaling, we study a generalization of entropy rates and extend the notion to generalized relative entropy rates.

When the functions involved in the generalized entropies are polylogarithmic functions and when the random sequence satisfy what we call "a local quasi-power property", we obtain explicit values for the entropy rates. Note that a wide class of finite or countable Markov chains satisfies this property as well as a wide class of random sequences with infinite correlations. Our results thus extend previous works of Rached et

al. (1999, 2001, 2004) that only deal with the classical Shannon or Rényi entropy rates with finite Markov chains.

Keywords: generalized entropies, entropy rates, Markov chains.

Discrete Time Homogeneous Markov Processes for the Study of the Basic Risk Processes

**Fulvio Gismondi[^], Jacques Janssen^{*}, Raimondo Manca[°],
Giuseppina Ventura**

[^]University "Guglielmo Marconi", Italy, ^{*} Solvay Business School
Université Libre de Bruxelles, Belgium, [°] Dipartimento di Metodi e
modelli per l'Economia, il Territorio e la Finanza, Università di Roma "La
Sapienza", ITALY

In this paper Markov models useful for following the time evolution of the aggregate claim amount and the claim number in the homogeneous time environment is presented. After the introductory arguments, the paper in the second section presents a short introduction to homogeneous discrete time Markov reward processes with discrete state number. In a third section it is explained how it is possible to follow the time evolution of the aggregate claim amount by means of a discrete time Markov process with a denumerable state number in homogeneous environment. The following step given in the paper is the introduction of rewards in the model. This is the fundamental step because in this way it is possible to study the time evolution of the system and the mean present value of the total amount of claims. Indeed, by Markov solution the mass probability function is reconstructed for each period and by means of reward processes it is possible to know the mean of the claim reserve for each year of forecasting. After it will be shown how it is possible to study also the mean total number of the claims.

Keywords: risk processes, Markov chains, reward processes, homogeneity & non-homogeneity

Nonparametric Algorithms of Identification, Forecasting and Control of Economic Systems

Irina L. Foox¹, Irina Yu. Glukhova², and Gennady M. Koshkin³

¹ Department of Computer Science, Tomsk State University, 36, Lenin
ave., 634050 Tomsk, Russia

^{2,3} Department of Applied Mathematics and Cybernetics, Tomsk State
University, 36, Lenin ave., 634050 Tomsk, Russia

Models of ARX-processes are widely used on identification of economic systems and financial time series analysis [1]. By identifying a model we

mean the problem of parametric or nonparametric estimation of the function defining ARX-process (see [2]-[4]).

The principal parts of mean square errors for kernel plug-in estimators of the functions defining ARX-process are found. We use simulation to compare parametric and nonparametric identification algorithms and to study also nonparametric control algorithms. To investigate the dependence of Russian Federation's Industrial Production Index on the dollar exchange rate, direct investments, and export for the period from September 1994 to January 2013, the proposed algorithms of identification and forecasting are applied.

Keywords: Kernel plug-in estimator, mean square error, ARX-process, nonparametric identification, forecasting algorithm, control.

Acknowledgement. This investigation was supported by the Russian Foundation for Basic Research (project no. 13-08-00744).

References

1. E. Masry and D. Tjøstheim. Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory*, 11, 2, 258-289, 1995.
2. A.V. Kitaeva and G.M. Koshkin. Semi-recursive nonparametric identification in the general sense of a nonlinear heteroscedastic autoregression. *Automation and Remote Control*, 2, 92-111, 2010.
3. A.V. Dobrovidov, G.M. Koshkin and V.A. Vasiliev. *Non-Parametric State Space Models*. Kendrick Press, Inc., Heber, UT 84032, USA, 2012.
4. G.M. Koshkin and I.Yu. Glukhova. Nonparametric identification of nonlinear ARX-processes. *Journal of Control and Computer Science. Tomsk State University*, 3, 20, 55-61, 2012 (in Russian).

Stochastic Processes in Nominal Exchange Rate: Out-of-Sample Forecasting Accuracy of the EUR/USD

Giray Gozgor

Department of International Trade and Business, Dogus University,
Turkey

It is well known that a random walk forecast of nominal exchange rates better than any fundamental model, so called as the exchange rate disconnect puzzle. The aim of this paper is to investigate the out-of-sample forecasting accuracy of the EUR/USD exchange rate over the period March 1999-March 2013. We compute Brownian Motion (BM), Geometric Brownian Motion (GBM), Discrete Brownian Motion (DBM), Ornstein-Uhlenbeck Mean-reversion (OUM), and Jump Diffusion (JD)

stochastic processes. We also consider the Vector Autoregressive (VAR), the Autoregressive Integrated Moving Average (ARIMA) models and the Taylor Rule fundamentals. We compare their forecasts in monthly and quarterly horizons, and find that only GBM and BM processes beat the Random Walk model. Furthermore, the results are robust and not time-specific. When we separately examine the periods of the pre-global recession and the post-global recession, the results remain unchanged.

Keywords: Exchange Rate Forecasting; Brownian Motion; Ornstein-Uhlenbeck; Jump Diffusion; Volatility Models; Probability Distribution

A Study on European Football Championships in the GLMM Framework with an Emphasis on UEFA Champions League Experience

Andreas Groll¹, Jasmin Abedieh

¹Department of Mathematics, Workgroup Financial Mathematics,
Ludwig-Maximilians-University, Germany

In this talk we address two major objectives. First, the results of a preceding article (Groll and Abedieh [1]) are revised, where all matches of the European football championship (EURO) 2012 have been predicted on the quite small data basis of the two preceding EUROS, resulting in a possible course of the tournament.

There, a pairwise Poisson model for the number of goals scored by national teams competing in EURO matches was established, incorporating two approaches for variable selection, which was then used for prediction. Including the data of the EURO 2012, in the present talk this analysis is replicated on a more reliable data basis and the set of selected influence variables is compared to the results of the preceding analysis. Besides, the course of the EURO 2012 suggests a positive correlation between a national team's success at a EURO and the number of its players that have been successful in the preceding Union of European Football Associations (UEFA) Champions League (CL) season. Hence, a second objective of this talk is to investigate, if in fact a significant influence of this covariate can be detected.

Keywords: Football, European football championships, UEFA Champions League, Sports tournaments, Generalized linear mixed model, Lasso, Variable selection.

References

[1] A. Groll and J. Abedieh. Spain retains its title and sets a new record - generalized linear mixed models on European football championships. *Journal of Quantitative Analysis in Sports*, 2013. To appear.

Inference for the GNL Distribution and Applications in Finance

I. Groparu-Cojocaru, L.G. Doray

Université de Montréal, Dept of Math. and Stat, CANADA

The generalized normal Laplace distribution has been used in financial modeling because of its skewness and excess kurtosis. To estimate its parameters, we use a method based on minimizing the quadratic distance between the real and imaginary parts of the empirical and theoretical characteristic functions. The quadratic distance estimator (QDE) obtained is shown to be robust, consistent and with an asymptotically normal distribution. The goodness-of-fit test statistics presented follow an asymptotic chi-square distribution. The performance of the QDE is illustrated by simulation results and an application to financial data.

Stationarity and Geometric Ergodicity of a class of Periodic Threshold ARCH Models

Hafida Guerbyenne, Abderrahim Kessira

Faculté de Mathématiques, Université des Sciences et de la Technologie Houari Boumediene (USTHB), Algérie

Since the paper of Yule (1927) on AR modeling of sun spots series until Box & Jenkins (1970) which marks the maturity of linear models in terms of theory and methodology, gaussian linear models prospered and dominated the theory explorations and the applications. However, Moran, in his paper on modeling Canadian Lynx series (Moran (1953)), showed the limitations of linear models explaining the different steps of population's fluctuation as "regime effect" (Tong (1990)). Modeling regime effect or other aspects non standard means going beyond linear gaussian models framework. This non standard aspects, which we call non linear aspects, include for instance non normality, asymmetries cycles, bimodality, time irreversibility, sensibility to initials conditions and others.

In this work, a class of periodic TARCH processes is introduced and studied. The existence of a strictly stationary and β -mixing solution is established under the positivity of the underlying independent white noise process density. We give sufficient conditions for the existence of moments. The analysis relies on Markov chain theory for a general state space.

Keywords: β -mixing, ergodicity, Markov chains, Periodic TARCH models.

Beta kernel regression estimator: Some comparisons with symmetric kernel

Zohra Guessoum¹, Sara Ghettab², Ourida Sadki³

¹Lab. M.S.T.D. Dept de Proba.-Stat., Fac. de Maths, U.S.T.H.B.,

²Département de Probabilités-Statistique, Faculté de Mathématiques, U.S.T.H.B., ³Université d'Oum El Bouaghi, Algérie

We investigate the asymptotic behavior of the non symmetric kernel regression estimator, in censored and independent model. Chen [3] and Bouezmani et al [1] introduced and studied the beta kernel density estimator with known compact support $[0, 1]$, in order to remove the boundary bias of the standard kernel introduced by Rosenblatt. We compare by simulation the performance of the beta kernel regression estimator with the symmetric kernel, performed by Guessoum and Ould-Said [4].

Keywords: kernel estimator, nonparametric regression, survival data, censored model, beta kernel.

A consistent estimation of the variance in a linear model with severe heteroscedasticity

Ali S. Hadi, Beatriz Lacruz, Ana Pérez-Palomares

University of Zaragoza, Spain

Estimation in parametric heteroscedastic linear models has been extensively studied during the last decades. Different methods have been developed and proposed in the literature to deal with different forms of heteroscedasticity. For example, the most common types of heteroscedasticity include the polynomial variance, multiplicative heteroscedasticity and the variances depending on the expectation of the model.

In this paper we deal with an alternative type of heteroscedasticity which provides unbounded variances. This kind of heteroscedasticity can produce inconsistency of the OLS estimation and therefore some of the methods developed in the literature would not give the desired results. The objective is to estimate the parameters involved in the variance in a consistent way and to construct an estimator of the expectation of the model which can be used in inference procedures. Simulation experiments are carried out to investigate the performance of the proposed method and to compare it with other known methods developed in the literature.

Keywords: Linear model, heteroscedasticity, unbounded variance, consistency.

Neural network modeling and fractional model identification of a thermal process

Sofiane Hammouche, Rabah Mellah, Tounsia Djamah, Rachid Mansouri

Laboratoire de Conception et Conduite des Systemes de Production (L2CSP), University M. Mammeri of Tizi Ouzou, Algeria

Modeling and identification of a complex dynamical system is an important task for its simulation, control design, prediction and fault diagnosis. Various non linear model structures have been proposed in the literature such as Voltera series, Wiener and Hammerstein models. On the other hand, diffusive phenomena exhibit complex dynamics that can be described by fractional models. These last are based on fractional derivative, known for its ability to take into account the memory effect; a typical example is the heat transfer process. Its is characterized by the hereditary property and infinite dimensional structure and the use of classical integer models is inappropriate. Another major limitation evolves, when temperature ranges in a wide interval, causing that certain thermo-physical parameters of the medium (conductivity and diffusivity) to vary and depend on the temperature. In this context, the process internal behaviour contains unknown non-linearities, difficult to model and not fully understood, and the use of intelligent computational methods such as neural networks may be attractive. They offer the advantages of the learning and adaptation mechanisms to adjust the non-linear model parameters, while conventional approaches tend to fail or become cumbersome.

The aim of this paper is the modeling and identification of the thermal process. To achieve this goal, two models are investigated; the first one is based on parsimonious fractional models, with at least two fractional orders to composite with the effects of the varying parameters with the frequency band, and the second one uses a multilayer neural network (NN) to reproduce the thermal process behaviour.

Numerical simulations on a thermal benchmark are performed in order to analyze the two models fitting ability to approximate the diffusive phenomena.

Keywords: Modeling, identification, heat transfer, difusive process, fractional order system, neural network.

Random Complex and Data Replication

Giseon Heo, Byron Schmuland, Yin Li

University of Alberta, Canada

The simplicial complex of a point cloud is constructed by connecting the points when the balls centered at the points intersect. Persistent homology studies the evolution of the topological features of a point cloud as the radius of the ball increases. Building the simplicial complex is deterministic since the nearest points are connected first and the second nearest point connected next and so on. We explore a different way of forming a complex by assigning probabilities to the edges of two points. The probabilistic modeling of complex can be viewed as a Markov chain. A current trend is for medical researchers to aim at developing personalized treatment for each patient. Our technique can be applied to compare a patient with a matched control or to monitor a patient's progress over time.

Keywords: Persistent homology, Betti numbers, Barcodes, Persistence diagrams, Persistence landscapes, Markov chains.

Making decision under uncertainty and Sport-Fantasy. An approach

Javier J. Hernández G.¹, María J. García G.², José G. Hernández R.³

¹Gerencia de Investigación, Minimax Consultores, UK, ²Gerencia General, Minimax Consultores, Venezuela, ³Departamento de Gestión de la tecnología, Universidad Metropolitana, Venezuela

In the sports business every day the fantasy games take a major relevancy, which enthuse a large quantity of fanatics. Among the problems that a fantasy participant has to is his squad of players. The conformation of this squad is usually very complicated in sports such as soccer, which besides the restrictions of budget there is also restrictions by positions. If it is taken into account, only the performance for the fantasy, of a determined player, the performance expected of that player for the next season it can be seen like a under uncertainty making decision problem. By which a hierarchy of the better players it is possible to obtain making use of a model of decision making under uncertainty, in particular The Amplitude Model (TAM). On how to use this model raises the objective of this work: Show how you can create a squad of players to participate in a fantasy game, based on The Amplitude Model.

Keywords: Sport, Fantasy, Soccer, Making decision under uncertainty, The Amplitude Model (TAM)

Modal Age at Death: Lifespan Indicator in the Era of Longevity Extension

Shiro Horiuchi¹, Nadine Ouellette², Siu Lan Karen Cheung³, and Jean-Marie Robine⁴

¹Hunter College, City University of New York (CUNY), USA, ²University of California, Berkeley, USA, ³University of Hong Kong, Hong Kong,

⁴Institut National de la Santé et de la Recherche Médicale (INSERM), France

This paper examines and demonstrates importance of the modal age at adult death (M) in the measurement and analysis of longevity in human populations. First, among three major indicators of central tendency of age distribution of deaths, the mean age (life expectancy at birth) and median age could be strongly affected by mortality of infants, children and young adults, but M is determined solely by old-age mortality, as far as the mortality schedule follows the regular “bathtub” pattern. Also M can be characterized as an indicator of the location of “death heap” in old age. Secondly, it can be shown that although conditional life expectancies at selected younger old ages such as $e(65)$ are widely used as a summary measure of old-age survival, they tend to fail to capture shifts of mortality pattern to older ages, which can be accurately captured by M . Thirdly, it is also shown that longevity differentials (e.g., by gender, country, region, ethnicity) are more properly reflected in differences in M than differences in other widely-used indicators of mortality and lifespan. Fourthly, M plays central roles in major mathematical models of age patterns of adult mortality such as Gompertz, logistic, Weibull models and their Makeham variants. Making use of the important fact that M is the age at which the force of mortality is equal to the life-table aging rate (the rate of relative mortality increase with age), it can be shown that those models are re-expressed using M as a key parameter in manners that are more clearly interpretable than the original forms of the models. Fifthly, although M may not be directly determined from erratic mortality data, the recently developed method for deriving M from the P-spline-smoothed mortality curve based on Poisson likelihood is highly effective in estimating M . The method thoroughly exploits observed age-specific death counts and population exposures data, as it does not impose any rigid assumptions about their functional form. In sum, this paper shows, with mathematical models and proofs as well as empirical examples, that M is an important and useful summary indicator of longevity. We claim that it should be used more widely and frequently in longevity research, with a recommended estimation method.

Keywords: modal age at death, longevity, mortality, models of age patterns of adult mortality, P-spline smoothing

Comparing small and large scale models of multicategory buying behavior

Harald Hruschka

University of Regensburg, Germany

We introduce a finite mixture of multivariate Tobit-2 models (FM MVT-2) to analyze multicategory decisions of households based on two response variables, purchase incidence and expenditure. In the empirical study we apply the FM MVT-2 model to 28 product categories which is considerably higher than the number of categories to be found in most previous publications on cross category dependences. By using expenditure instead of quantity we are not forced to make assumptions about equivalences of quantities of different product forms or different qualities which belong to the same category. Obtaining reasonable assumptions on such equivalences becomes the more difficult the more categories one has to analyze. Another advantage of expenditures is the fact that compared to quantities they are more directly related to managerial objectives like sales revenue and gross margins.

Our use of a finite mixture approach can be justified by statistical performance and managerial usefulness.

The finite mixture approach allows us to identify cross category dependences which are segment specific. In our empirical study the number of cross category dependences turns out to differ between segments.

A literature overview of publications on cross category dependences shows that typically only few categories are considered. Authors of these publications assume that buying behavior can be described with sufficient accuracy even if attention is limited to dependences among few categories. But looking at a small number of categories could entail several problems. Firstly, because other categories are ignored it could lead to an inaccurate description of buying behavior. Secondly, even households' decisions on the few categories which are considered could depend on decisions on categories which are left out. Thirdly, any of the small scale models could lead to biased results by indicating cross category dependences which do not turn out in the large scale model. Fourthly, managers who base decisions on such models will underestimate marketing effects if categories which are analyzed in different small scale models are positively correlated.

We investigate whether these problems occur and, given that they do, how important they are. To this end we compare the best performing comprehensive 28 categories FM MVT-2 model to several small scale models each of which considers seven categories. We want to determine whether the forecasting performance of the four small scale models taken together is worse than that of the large scale model. We also want

to know if results on cross category dependences differ between the the four small scale models and the large scale model.

Keywords: Multicategory buying decisions, cross category dependences, Tobit model, finite mixture, Markov chain Monte Carlo estimation

Binary dynamic response model with structural changes

Šárka Hudecová

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic

Binary time series play an important role in many fields of application, including macroeconomics and financial econometrics.

A common approach to modelling a binary series $\{y_t\}$ is based on a *dynamic probit model* (also *binary dynamic response model*). It is assumed that the probability of a success at time t , denoted as π_t , depends on lagged values of the series $\{y_t\}$ and on explanatory variables X_t via a probit function

$$\Phi^{-1}(\pi_t) = \sum_{j=1}^p \gamma_j y_{t-j} + \beta' X_t,$$

where Φ is the CDF of the standard normal distribution, γ_j and β are unknown parameters, and $p \geq 1$.

However, in practice, the data-generating mechanism may change at some unknown time point (change point). The main objective is then to decide whether a change has occurred and to detect the change point.

We study dynamic binary response models with possible changes in their parameters. A cumulative sum (CUSUM) type test statistic is proposed for testing a change. The asymptotic distribution of the test statistic is derived under the null hypothesis. The performance of the test is illustrated by a simulation study.

Keywords: binary time series, binary dynamic response model, dynamic probit model, change point analysis

Cascade Models for Random Financial Networks

Tom Hurd

Department of Mathematics and Statistics, McMaster University, Canada

In 2001, Eisenberg and Noe put forward a simple accounting framework for analyzing default cascades in a deterministic financial network. Subsequent work has pushed the framework in many directions, notably in making more and more of its parts stochastic to reflect the numerous sources of uncertainty in our observations of real world networks. After reviewing some of the important variations that motivate the development of new mathematical techniques, I focus on a classic model from network science, the Watts 2002 Cascade Model. The basic structure of this threshold model is so universal that it has been used to understand such diverse phenomena as disease transmission, the spread of rumours and fads, the evolution of species, the collapse of power systems and avalanches in sandpile models. Its underlying simplicity will allow me to spend the remainder of the talk exposing some of the beautiful mathematical properties that lie buried within.

Keywords: Contagion, random graph, stochastic network, information cascade, bond percolation, systemic risk, banking network.

Truncation Approximations of the M/G/1 Queue: Strong Stability Approach

Badredine Issaadi¹, Karim Abbas², Djamil Aïssani²

¹Department of Mathematics, Sciences Faculty, University M'hamed Bougara of Boumerdes, ²Laboratory of Modeling and Optimization of Systems (LAMOS), Exact Sciences Faculty, University of Bejaïa, Algeria

Queueing models are well known for modeling and evaluation purposes of real systems from areas as diverse as telecommunications, manufacturing and servicing. Particularly, the M/G/1 queue is very used to model many practical applications. For computing the steady-state distribution of the M/G/1 with infinite capacity, we use generally a numerical method to invert the Laplace-Stieltjes transform of the distribution. For this reason and when we deal with infinite state space of this kind of systems, we must suggest to modify the structure of the system or to consider the finite version of the capacity of the same system. In this way, many truncation approaches are available to approximate the performance measures of the infinite state space of such models by that of the truncated one.

In this talk, we follow a different train of thought, and will present easy computable new bounds, by using the strong stability approach, on the effect for switching from the infinite to finite capacity to predict the

performance measures of the M/G/1 queue. We also consider in our analysis two different approaches to truncate the space state of the model. In addition, we illustrate with some numerical examples the resulting approximations turn out to be of practical value.

Keywords: M/G/1 queue, Augmented truncation, Approximations, Performance measures.

TO RELIABILITY OF MORTALITY SHIFTS IN WORKING POPULATION IN RUSSIA

**Alla Ivanova, Tamara Sabgayda, Viktoria Semyonova, Elena
Zemlyanova**

Department of Health Statistics, Federal Research Institute for Health
Organization and Informatics of Ministry of Health of the Russian
Federation, Russia

The article presents the analysis of death causes structure of working population (15-59 years) in Russia determining its age-specific and gender regularities and defining their influence on nosological profile of mortality in working ages. The main results of 50-years of history of Russian mortality are the following:

- up to the end of the first decade of XXI century Russia returned to 45-years ago life expectancy levels; this visible reversion constitutes of women's gain of 1.3 years and the equal disadvantage in men;
- the main source of both disadvantages and gains during all stages of nearly half century dynamics is the population of working ages; in males cumulative losses due to population of 15-59 years following the results of 1965-2010 occurred to be 2.5 years of life expectancy, in females – 0.5 years.
- in the ages where resulting mortality during analyzed period didn't change in general, and in ages where it increased the structure of death causes visibly changed:
- as to structure of mortality in young ages, input of respiratory and digestive diseases as well as ill-defined conditions plus infections and cardio-vascular diseases in males the resulting mortality remained the same as in 1965 only due to reduction of traumas and poisonings and neoplasms;
- as to structure of mortality in working ages over 30 years, input of main somatic diseases (except neoplasms) and external causes increased which determined growth of summarized mortality in these ages.

Keywords. Mortality, input of age groups, input of separate death causes, working population, reliability of mortality data

Time aspects of a fund manager appraisal

Evgeny A. Ivin, Alexey N. Kurbatskiy, Alexandr V. Slovesnov

Moscow School of Economics of Moscow State Lomonosov University,
Russia

We deal with a problem of estimating the performance of management of a unit investment trust. If we speak about the Russian equity market the estimation is usually based on comparison of fund's rates with appropriate characteristics of the Russian benchmark that is the MICEX index. Every manager fills his fund with investment ideas but he is able to estimate only approximately the probability with which they will be realized. It may turn out that some of his ideas won't be realized at all or they will be affected by unpredictable scenarios. In the paper we estimate the time necessary to make a conclusion about manager's performance and show that short-term periods are not optimal a priori and it's not worth to invest in Russian stocks for only a few months.

Comparing yields of the leading Russian managed fund with the MICEX index we use different approaches to estimate necessary time for reliable conclusion about the effectiveness of the fund manager. It is also shown that the MICEX index can't be considered as a benchmark to compare different funds and such universal index is suggested.

Keywords: yield, volatility, equity market, fund manager, weighted index, Hurst coefficient, martingale, financial time series

Calculation of semi-Markov flow characteristics

Ivnitskiy V.A.

Moscow State University of Railway Transport, Russia

Article is devoted to finding arbitrary order initial moments and a correlation functions of the process $\xi(t)$, meaning a time interval to the next time arrival, and of the process $\nu(t)$, meaning a number of arrivals before a time t , for semi-Markov flow under general initial conditions.

Keywords and phrases: semi-Markov flow, semi-Markov process, arbitrary order initial moments, correlation functions, Laplace transformations, general initial conditions.

Age-Period-Cohort models using smoothing splines: a generalized additive model approach

Bei Jiang, KC Chough

Dept Mathematical and Statistical Sciences, University of Alberta,
Canada

Age-Period-Cohort (APC) models are to analyze temporal trends in disease or mortality rates, dealing with linear dependency among the associated effects. However, its use has been severely limited due to the nature of sparse data. To deal with these practical limitations and issues, we advocate cubic smoothing splines to smooth the effects of age, period and cohort. We show that the methods of estimable functions proposed in the framework of Generalized Linear Models can still be considered a solution to the non-identifiability problem when the model fitting is within the framework of Generalized Additive Models with cubic smoothing splines. Through simulation studies, we evaluate the performance of the cubic smoothing splines in terms of the MSEs of estimable functions.

Our results support the use of cubic smoothing splines for APC modeling with sparse but unaggregated data from the Lexis diagram.

Keywords: Age-Period-Cohort models; Generalized Additive Models; smoothing splines; estimable functions

Methods and Techniques for Multifractal Spectrum Estimation in Financial Time Series

Petr Jizba, Jan Korbel

FNSPE, Czech Technical University in Prague, Czech Republic

In this paper, we compare two key approaches used in time series analysis, namely the Multifractal Detrended Fluctuation Analysis and Multifractal Diffusion Entropy Analysis. The comparison is done from both theoretical and numerical point of view. To put some flesh on bare bones, we illustrate our analysis by applying both methods to three model time series. As a fourth illustration we analyze the empirical time series of daily returns of S&P500 stock index recorded over the 50 years period. We argue that while the Multifractal Detrended Analysis is computationally more efficient, the Multifractal Diffusion Entropy Analysis is conceptually cleaner. In addition, the latter allows a wider applicability in cases when time series have underlying distributions that are heavy tailed.

Keywords: Multifractal spectrum, Detrended Fluctuation Analysis, Rényi Entropy.

Estimation of Sasang Constitutional Distribution using Direct Standardization Method

Hee-Jeong Jin, Young Hwa Baek, Ho-Seok Kim, Si-woo Lee

KM Health Technology Research Group, Korea Institute of Oriental Medicine, Korea

Recently, there are many studies have produced interesting individual's differences by Sasang constitution (SC). These differences by SC are used to overcome human's diseases and to improve the quality of health care.

Sasang Constitutional Medicine (SCM) was established by Jema Lee under the Korean medical tradition. It classifies humans into four constitution types, which are Taeyangyin (TY), Tae-eumin (TE), Soyangyin (SY), and Soeumin (SE), depending on the nature of an individual's physiological, psychological, and physical characteristics that lead to differential responses to herbs. Until now, most of SCM studies have focused on these points- sensibility about individual herbal medicine. Though many researchers have studied based on SC, the distributional rate of SC has not changed since 1900.

In this study, we estimated the distribution of SC in Korea using Korea Constitution Multicenter Study subjects applying the direct standardization. After surveying 24 Korean medicine clinics and universities, 3,711 subjects were recruited and clinical information was collected. We applied the direct standardization method with the standard population from '2010 Korean Population and Housing Census'. The estimated constitutional distributions were TE 39.2%, SE 27.1%, SY 33.7%. Estimate the distribution rate of SC will be used an important factor in deciding health policy or developing new drug reflecting the results of previous SCM studies.

Keywords: Sasang constitutional medicine, Sasang constitution, direct standardization method, distributional rate.

Prevalence projections of chronic diseases and impact of public health intervention on risk factor: application on dementia in France

Pierre Joly

Isped, Université Bordeaux Segalen, France

Dementia is a frequent disease in the elderly population and a major public health problem. The number of cases (prevalence) of dementia will increase in the future, mainly because of the increase of life

expectancy. The prevalence is the number of cases of a disease in a population at a specific time and it can be important information to plan future actions of the health system. In general, prevalence is estimated from cross-sectional studies on representative samples of the population while the estimation of the incidence requires longitudinal studies. We use demographic projections for general mortality and population sizes. Age-specific incidence of dementia and age-specific mortality of demented subjects can be estimated from a cohort study. From these quantities we propose how to estimate the age-specific prevalence of dementia. We do not make the hypothesis that the mortality of non-demented subjects is equal to the general mortality. Furthermore, we develop an approach to estimate changes in future disease prevalence resulting from prevention campaign to reduce the frequency or the excess risk associated with a risk factor.

Spatial Variation in the Relationship between Mortality Rates and Neighborhood Characteristics in South Korea

Yoohyung Joo, Hee Yeon Lee

Department of Environmental Planning, Graduate School of
Environmental Studies, Seoul National University, Korea

This paper study examines spatial patterns of standardized mortality rates in South Korea and examines whether there are “spatially heterogeneous effects” of neighborhood context on mortality rates. To examine the spatial variation of neighborhood impact, this research uses geographically weighted regressions (GWR). The dependent variable is standardized mortality rates and the explanatory variables are neighborhood characteristics. They are categorized into four sectors: economic condition, social capital, housing environment, and built environment. This study uses representative indicator of each sector. The study finds that the relationship between neighborhood environments and mortality rates is spatially heterogeneous across geographical areas. This indicates that the global regression model fails to account for the spatial variation in the association between mortality rates and neighborhood environments.

The Spatial Distribution of the South American Sea Lion along the Central Coast of Chile

Bent Jørgensen¹, Ronny Vallejos², Felipe Osorio²

¹Department of Mathematics and Computer Science, University of Southern Denmark, Denmark, ²Department of Mathematics, Universidad Técnica Federico Santa María, Chile

We consider aerial survey data [2] on the spatial distribution of sea lions along the central coast of Chile collected for four years between 1998 and 2012. Like many other data on the spatial distribution of animal populations, these data present a serious challenge to conventional geostatistical methods because of the so-called proportional effect, i.e. non-constant spatial variance. We show that the data satisfy a power law for the spatial variance as a function of the spatial mean and the size of the surveyed area. This is consistent with a self-similar Tweedie spatial process with equicorrelated covariance structure and long-range dependence, following Jørgensen et al.[1]. The model is fitted using quasi-likelihood for the regression parameters and Pearson estimating functions for the covariance parameters.

Keywords: Geostatistics, Long-range dependence, Power variance function, Proportional effect, Scale invariance, Self-similarity, Taylor's power law, Tweedie spatial process.

References

1. B. Jørgensen, J.R. Martínez and C.G.B. Demétrio. Self-similarity and Lamperti convergence for families of stochastic processes. *Lith. Math. J.*, 51, 342-361, 2011.
2. M. Sepúlveda, D. Oliva, A. Urra, M. J. Pérez-Álvarez, R. Moraga, D. Schrader, P. Inostroza, Á. Melo, H. Díaz, and W. Sielfeld. Distribution and abundance of the South American sea lion *Otaria avescens* (Carnivora: Otariidae) along the central coast off Chile. *Revista Chilena de Historia Natural*, 84, 97-106, 2011.

Exit from a semi-stochastic interval by a Brownian motion

Tetyana Kadankova

Vrije Universiteit Brussel, Department of Mathematics, Belgium

We obtain the explicit formulae for the joint distribution of the first crossing time of two Brownian motions and the position of the processes at this moment. Further we derive the joint distribution of the first exit time from the semi-stochastic interval by the Brownian motion and the position of the process at this instant. Based on these results, we

calculate the expected value of the first exit time from the semi-stochastic interval and obtain the asymptotic expansions for the distribution of the first exit time from the semi-stochastic interval. We discuss briefly possible applications of these results in Biology and Physics.

Queuing Systems with Two Service Operations as Mathematical Models of Reliability and Survivability

Revaz Kakubava¹, Tinatin Kaishauri², Otar Shonia²

¹Department of Applied Mathematics, Georgian Technical University,

²Management Information Systems, Georgian Technical University,
Georgia

The given paper deals with the redundancy and maintenance problem for a wide class of any territorially distributed standby systems consisting of unreliable repairable elements. Mathematical models for interaction of degradation and its compensation processes in the above mentioned systems are proposed and their possible applications are partially analysed. These models represent mixed type queuing systems for two parallel maintenance operations – replacements and repairs. The problem for optimization of said system by economic criterion is stated. The possible ways of its solution are discussed.

Keywords: queuing models; structural control; maintenance; replacement; renewal (repair).

Coxian Phase-type distributions for studying students' length of graduation: Experience from an Italian and a Greek University

Aglaia Kalamatianou¹, Adele H. Marshall², and Mariangela Zenga³

¹ Department of Sociology, Panteion University of Athens, c/o 32 Alexandrou Panagoulis, Glyfada Attikis 16675 Greece

² Centre for Statistical Science and Operational Research (CenSSOR) Queen's University of Belfast, Belfast, BT7 1NNA, UK

³ Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via B. degli Arcimboldi, 8, 20126 Milano, Italy

Italian and Greek university systems show similar rules for the graduation: there is a time threshold for graduation, but students can graduate any time after this threshold without a time limit. This means that the duration of graduation could last for a long time for some group of students.

The aim of this paper is to analyse students' progression to graduation to estimate the influence of various factors on the probability that students, with certain characteristics, will progress successfully towards their degree or be still enrolled at the end of the observation. We propose to use the Coxian phase-type distributions for modelling the length of graduation of the students enrolled at an Italian and a Greek public university.

Keywords: duration of studies, lifetime-type educational data, survival analysis, Coxian Phase type distribution.

Measuring Credit Risk of Individual Corporate Bonds and Deriving Term Structures of Default Probabilities

Takeaki KARIYA, Yoshiro YAMAMURA, Koji INUI, Zhu WANG

Meiji University, Graduate School of Global Business, Japan

No doubt, the importance of empirical credit risk analysis has ever been increasing not only in financial industries but also in business and even in government under increasing world-wide uncertainties. In this paper, basing our arguments on the model of pricing government bonds (GBs) in Kariya et al (2012) and using corporate bond (CB) prices as our data source on credit, we first propose a measure of credit risk price spread (CRPS) for each CB relative to a GB-equivalent CB price. To choose an empirically effective CRPS measure, we test a hypothesis of no attribute preference with respect to investors' behaviors forming prices in the market of GBs. The results strongly reject the hypothesis against maturity preference as well as coupon preference. Secondly using the CRPS measure, a specific agency credit rating is shown to be ineffective for making credit-homogeneous groups of corporate bonds, where industry category is also used. To get our credit-homogeneous grouping, the CRPS measure is standardized by adjusting the differences of maturities and a three-stage cluster analysis is applied to the observed standardized CRPSs for Japanese CBs to get 14 groups, where 1545 CB prices as of 2010.8 are included. Since the grouping by the cluster analysis is a posterior grouping which is based on stochastically realized CBs and GBs, we propose Fixed Interval Rating (FIR) Method based on the standardized CRPS, and form 10 credit homogeneous groups. Thirdly, we derive the term structures of default probabilities (TSDPs) for some cluster groups and FIR groups and some individual firms via Kariya (2012) model, where industry factor is also considered. Naturally the TSDPs reflect the investors' future perspective on defaults of individual firms or groups.

Keywords: credit risk price spread (CRPS), corporate bond, cluster analysis, term structure of default probabilities (TSDP), investors'

behavior in the government bond market, default intensity model, fixed interval rating (FIR)

A Robust Approach To CoPlot Analysis

Yasemin Kayhan Atilgan, Suleyman Gunay

Department of Statistics, Faculty of Science, Hacettepe University,
Turkey

Coplot method is a useful graphical display technique which enables the simultaneous investigation of the relations between the observations and between the variables for a set of multivariate data. The aim of this method is to obtain a simple picture of the data, which gives opportunity to the researcher to make richer interpretation. At CoPlot analysis multidimensional scaling and Pearson's correlation coefficient are used to obtain a map which demonstrates observations and variables simultaneously. However, both multidimensional scaling and Pearson's correlation coefficient are very sensitive to outliers. When data set contains even if strong one outlier the map produced by CoPlot may represent the wrong relationships. In this study it is recommended to use robust multidimensional scaling and robust correlation coefficient in order to eliminate the negative effects of outliers upon the CoPlot map. By giving numerical examples the merits of using robust techniques in CoPlot analysis are demonstrated.

Keywords: MAD correlation coefficient, Robust Multidimensional scaling, Multivariate data analysis, Outliers.

Retrial queues with disruptive and non-disruptive server interruptions

TEWFIK KERNANE

Department of probability and statistics, Faculty of Mathematics,
University of Sciences and Technology USTHB, Algeria

We consider a single server retrial queue with the server subject to interruptions with constant and classical retrial policies for the access from the orbit to the server. We analyze the equilibrium distribution of the system and obtain the generating functions of the limiting distribution. Some performance characteristics are given such as mean queue lengths in orbit and in the system and the mean waiting time in steady state.

Keywords: retrial queue; disruptive interruption; non-disruptive interruption; steady-state distribution.

Robustness in Sequential Hypotheses Testing Applied to Statistical Monitoring of Medical Data

Alexey Kharin

Department of Probability Theory and Mathematical Statistics,
Belarusian State University, Belarus

Sequential approach to the problem of statistical hypotheses testing is an effective technique of data analysis (Wald[1]). It minimizes the expected number of observations provided the error probabilities are less than the maximal possible levels given (Mukhopadhyay and de Silva[2]). This optimality property proved by the theory is often broken in applied problems as real data do not follow a hypothetical model exactly, in other words, the hypothetical model is distorted (Huber[3], Kharin and Shlyk[4], Kharin[5]), and the performance characteristics of sequential tests (error probabilities values, expected sample sizes) increase significantly (Kharin[6]).

The problem of robustness of sequential tests is analyzed theoretically for simple and composite hypotheses cases. The following types of distortions are considered: "outliers" in observations, "contamination" of prior probability distributions of parameters, neighborhoods in L1- and C-metrics. Asymptotic expansions for the test performance characteristics are constructed with respect to the distortion level value. With the use of main terms of these expansions, the deviations of the performance characteristics under distortions from the hypothetical values are evaluated.

A parametric family of robustified sequential tests is proposed, and within this family the robust sequential test is constructed by the total error probability maximal value minimization provided the expected sample size is constrained. The theory is applied to the problem of incidence data monitoring.

Keywords: Sequential Decision Rule, Distortion, Robust Test, Incidence Data

The research is supported by the project ISTC B-1910.

References

1. A. Wald. Sequential Analysis, Wiley, New York, 1947.
2. N. Mukhopadhyay and B. de Silva. Sequential Methods and their Applications, Chapman & Hall / CRC, London, 2009.
3. P.J. Huber. Robust Statistics. Theory and Methods, John Wiley and Sons, New York, 2004.
4. A.Yu. Kharin and P.A. Shlyk. Robust Multivariate Bayesian forecasting under functional distortions in the chi-square metric. Journal of Statistical Planning and Inference, 139, 3842-3846, 2009.

5. A.Yu. Kharin. Robust Bayesian prediction under distortions of prior and conditional distributions. *Journal of Mathematical Sciences*, 126, 1, 992-997, 2005.
6. A. Kharin. Robustness analysis for Bayesian sequential testing of composite hypotheses under simultaneous distortion of priors and likelihoods. *Austrian Journal of Statistics*, 40, 65-74, 2011.

Diffusion Approximation in Inventory Management

Anna V. Kitaeva, Natalya V. Stepanova

National Research Tomsk Polytechnic University, Russia

Single-product inventory management model with both random and controllable demand and continuous input product flow with fixed uncontrolled rate under finite storage capacity is considered. It is shown that under some conditions we can consider the stock level process as asymptotically diffusion process and find its stationary distribution. The result permits us to solve the optimization problem and control the probabilities of the stock-out and overflow.

Keywords: Inventory Control, Stochastic Model, Diffusion Approximation.

Impact of dependencies on insolvency probabilities

Orsolya Kocsis

Hungarian Central Statistical Office, Hungary

Insurance companies always estimate the number and amount of future claims. Under the compound risk model, there is usually an assumed independence between these variables. Besides, there are situations when some degree of dependence is considered. In this paper, the dependence is assumed along with the risk profiles, which are derived from Sarmanov-Lee family of distribution. The model allows us to determine the number and aggregated amount of future claims with Bayesian analysis. For the estimation of the aggregated amount, two types of cases are presented: when risk profiles are the same and when they are different by contracts. The difference between the independent and dependent cases can be observed using simulated data. The simulation has been done in R program.

Keywords: Insurance, compound risk model, Bayes, dependence, Sarmanov

Skew t -copula and tail dependence: a simulation study

Tõnu Kollo¹, Gaida Pettere², Marju Valge¹

¹University of Tartu, Estonia, ²Riga Technical University, Latvia

Let X and Y be two (possibly dependent) random variables with marginal distribution functions $F(x)$ and $G(y)$. Then the joint distribution function $H(x,y)$ can be presented through a copula. In applications an important property of a copula is tail dependence. Tail dependence refers to the degree of dependence in the corner of the lower-left quadrant or upper-right quadrant of a bivariate distribution. Will a big value of X increase the probability that Y will also be big? An answer to this question is characterized mathematically by the limit of the conditional probability $P(Y|X)$ when X tends to infinity and Y is greater than or equal to X . The limit is the characteristic of tail dependence. t -copula takes into account tail dependence when using it as a data model. In 2010 the skew t -copula was introduced by T. Kollo and G. Pettere. The construction is based on the multivariate skew t -distribution. First applications have shown that the distribution can successfully be used in practice. To examine the tail behavior of the skew t -copula and speed of convergence to the theoretical value we have carried out a simulation experiment. Estimates by maximum likelihood and moments' method have been found and convergence to the theoretical value of the tail dependence coefficient studied in both cases and for several parameter sets.

Keywords: simulation method, skew t -copula skew t -distribution, tail dependence.

Algorithm to maximize the degree of attainability and the degree of desirability of manpower planning

Komarudin¹, Marie-Anne Guerry¹, Tim De Feyter², Greet Vanden Berghe^{3,4}

¹Vrije Universiteit Brussel, MOSI, Brussel, Belgium, ²Center for Business Management Research, Hogeschool-Universiteit Brussel, KU Leuven Association, Belgium, ³KAHO Sint-Lieven, CODES research group, Gent, Belgium, ⁴KU Leuven, Department of Computer Science, Kortrijk, Belgium

This work studies a Markov manpower planning problem concerned with fulfilling the long term personnel need. Two criteria are considered for evaluating a personnel structure, i.e. the degree of attainability and the degree of desirability. The degree of attainability is an extension of the concept attainability described by Bartholomew et al. (1991) to overcome

the crisp division of the set of the attainable and the set of the unattainable personnel structures.

The degree of attainability represents the similarity of a personnel structure and the set of personnel structures that are attainable (Guerry, 1999). The degree of desirability is the degree in which the personnel structure corresponds with the long term objectives defined by the top management.

The degree of desirability reflects the degree of similarity of a personnel structure and the desired personnel structure. In previous work, both criteria have been defined based on fuzzy set theory.

Current state of the art has combined these two similar criteria (De Feyter & Guerry, 2009). However, the discussion of the degree of attainability is always under the assumption that the personnel structure has a fixed value for the total size, i.e. the total number of personnel. At present no optimization method has been presented for finding a personnel structure with a high value for the degree of attainability as well as for the degree of desirability.

We extend the previous model and propose two new models that provide more flexibility.

The new models are flexible in terms of providing generalizations by, for example, allowing fixed total size constraint relaxation.

Moreover, the new models can accommodate additional constraints, such as a required ratio between the number of personnel in one subgroup to another, etc.

The two models are formulated using different membership functions.

The first model uses a reciprocal function to formulate the degree of attainability and the degree of desirability whereas the second model employs a triangular function to define them. The first model can be formulated as a linear program.

However, the second model can only be formulated as a mixed integer nonlinear program.

Two different algorithms are proposed to address the second model, i.e. piecewise linear approximation (PLA) and particle swarm optimization (PSO).

Extensive experiments have been performed to instances derived from existing nurse rostering data sets, in order to assess the algorithms. It is shown that each of the proposed algorithms can be used to effectively solve its respective model.

Keywords: personnel structure optimization, manpower planning, Markov model, degree of attainability, degree of desirability

REFERENCES

- David J. Bartholomew, Andrew F. Forbes, and Sally I. McClean. Statistical techniques for manpower planning. John Wiley & Sons, 1991.
- Tim De Feyter and Marie-Anne Guerry. Evaluating Recruitment Strategies Using Fuzzy Set Theory in Stochastic Manpower Planning. *Stochastic Analysis and Applications*, 27(6):1148-1162, 2009.

Marie-Anne Guerry. Using Fuzzy Sets in Manpower Planning. *Journal Applied Probability*, 36(1):155-162, 1999.

Modelling structural changes in relations between returns of selected stock indexes

Jozef Komornik¹, Magda Komornikova², Cuong Nguyen³

¹Faculty of Management, Comenius University, Slovakia, ²Faculty of Civil Engineering, Slovak University of Technology, Slovakia, ³Faculty of Commerce, Lincoln University NZ, New Zealand

We have investigated the relations between 8 selected countries' (USA, Canada, Japan, Australia, Hongkong, Singapore, UK and France) daily returns of stock indexes in the time period January 3, 2000 – May 8, 2012, divided in 3 subperiods bounded by the recent global financial crisis (July 1, 2008 – April 30, 2009). We have observed that in the postcrisis subperiod the influence of the delayed values (by 1 day) of the returns of the US stock market on the markets of Japan and Australia greatly increased, while the same effect did not take place for the remaining (even East Asian) indexes. We used the copula approach for fitting the optimal models for the investigated relations. We employ Archimedean and mixed copula models of Clayton, Gumbel, Frank and Joe families as well as their survival copulas. The flexibility of using mixed copulas is that it allows us to capture different structures among variables, including both right and/or left tail dependence structure at the same time.

Our results prove that using copulas, especially mixed copulas, to capture dependencies is a useful and flexible approach. There are several opportunities for future research, such as extending the number of stock markets and/or examining the relationship between different investments in commodity markets, bond and stock markets, using the copula approach for fitting the optimal models for the investigated relations (possibly also with an extended range of copula families, for example asymmetric logistic model copula (ALM), and mixed asymmetric logistic model copula (MALM)).

Keywords: Stock index, Returns of index investments, Copula, Archimedean copula, Reflection of copulas.

Acknowledgement

This work was partially supported by Slovak Research and Development Agency under contract No. APVV—0073--10 and by VEGA 1/0143/11.

Conditional Density Estimations

Katerina Konecna, Ivanka Horova, Jan Kolacek

Department of Mathematics and Statistics, Masaryk University,
Czech Republic

We consider the kernel estimator of conditional density which was introduced in [1]. Conditional density can be regarded as generalization of regression – while regression models conditional expectation, conditional density models the whole distribution – one can have better idea on both – the whole data visualization as well as the shape while supposing fixed observation X_i .

Smoothing parameters in x and y directions play a principal rule of quality of these estimates and control their smoothness. We consider diagonal smoothing matrix which minimizes MISE (Mean Integrated Square Error), the indicator of quality of estimates. We will focus on methods of known conditional density estimators, especially on Nadaraya-Watson estimator. We pay attention to methods of choosing the smoothing parameters. These methods are compared on simulated data as well as a real data set.

Keywords: Kernel smoothing, conditional density, bandwidth selection.

References

[1] R.J. Hyndman, D.M. Bashtannyk, G.K.Grunwald. Journal of Computational and Graphical Statistics, *Estimating and Visualizing Conditional Densities*, Volume 5, Issue 4, 1996

Optimizing properties of adaptive strategy in computing system with different workload

Mikhail G. Konovalov

THE INSTITUTE OF INFORMATICS PROBLEMS
OF THE RUSSIAN ACADEMY OF SCIENCES (IPI RAN), Russia

The effective job allocation policy can essentially improve the productivity of distributed computing system without accretion of computer power pool. Recent studies have shown the usefulness for this aim of adaptive flow control algorithms based on the theory of partially observed Markov chains [1, 2]. This report continues the analysis of relations between system performance and adaptive algorithms parameters. Special attention was paid to the question, how responds the adaptive strategy on the various workload. The model of the adaptive controlled distributed computing system has been investigated under different job flow types, including self-similar processes with long-range dependence. The results were obtained concerning optimal choice of strategy parameters.

1 Konovalov, M. G., Malashenko, Yu. E. and Nazarova, I. A., Job Control in Heterogeneous Computing Systems, *Journal of Computer and Systems Sciences International*, 50, 1, 43–61 (2011).

2 Konovalov, M. G., Computer system optimization using simulation and adaptive algorithms, *Informatics and its applications*, 6, 1, 37-48 (2012) [in russian].

Keywords: Markov decision process, adaptive algorithm, computing system, self-similar process

Reliability Evaluation of Multi-Camera Motion Detector by using Monte-Carlo Simulation

Samuel Kosolapov

Signal and Image Processing Laboratory, ORT Braude Academic
College of Engineering, Karmiel, Israel

Motion Detectors (MD) utilizing one digital camera, are well-known and widely used for detection of physical objects intrusion into protected zone. For most MD any significant change in the content of the frame grabbed by digital camera is treated as "security violation event". This means that MD operation is limited to the protection of 2D region of fixed size, which significantly limit practical usage of MD. Stereo Motion Detectors (SMD, or 3DMD – 3D Motion Detector) utilizing two video cameras, can detect physical violation of the user specified 3D volume, but, as it was shown by earlier research, two-camera setup has low reliability for some motion paths, which lower total SMD reliability. Constantly dropping prices on high resolution digital cameras makes practical implementation of Multi-Camera Motion Detector (MCMD), at least for the case of 3 or 4 digital cameras. It seems obvious that increase in the number of cameras increases MCMD reliability, however, not every multi-camera setup is practically feasible because of camera(s) calibration need. Operation of SMD and MCMD in most cases requires some kind of calibration, which, in some cases, is problematic in the real-life conditions. Hence, it would be preferable to utilize setups that can aligned during assembly, thus, effectively eliminating need for "after-assembly" calibration. In order to evaluate accuracy and reliability of the selected "aligned" MCMD setup, software simulator was designed and implemented by using MAPLE. Simulator takes into account camera' parameters and noise, effect of digitization and assembly errors of MCMD setup. User can select parameters and assembly tolerances of the MCMD setup, geometry of 3D volume to be protected, intrusion

object size and its motion path. First unit of this simulator generates plurality of digital images in the situation when 3D object of specified size and shape is moving over specified 3D path. Second unit (by processing simulator-generated or real images) is attempting to detect intrusion of this 3D object into user specified "protected volume". Well-known mathematical models of different 3D setups cannot be used directly to evaluate accuracy and reliability of MCMD, because a number of parameters cannot be measured exactly for non-calibrated setup. Thus, classical Monte-Carlo approach was used to evaluate accuracy and reliability of a number of practically interesting MCMD configurations. In the step of Monte-Carlo simulation, software unit is operated a number of times, while, for every simulation run, values of the selected set of setup parameters are modified in a pseudo-random way. Simulation results in evaluation of the "true positive", "true negative", "false positive" and "false negative" of the selected setup. Additionally, results of simulations can be organized as a set of 2D and 3D plots enabling to recommend customer-tailored MCMD configuration for the specified volume to be protected. Reliability of a number of practically interested setups were analyzed and compared.

Keywords: Image Processing, 3D Imaging, Stereo Camera, Motion Detector, Monte-Carlo simulation, MAPLE

A new principal component method to analyse frequency tables with different row and column sets using common instrumental variables: application to a multilingual survey

Belchin Kostov^{1,2}, Mónica Bécue-Bertaut², François Husson³

¹IDIBAPS, Spain, ²Universitat Politècnica de Catalunya, Spain

³Agrocampus-Ouest, France

We propose a new methodology to deal with closed and open-ended questions answered in different languages. For each language l ($l=1, \dots, L$), the answers are coded into a multiple table (X_l, Y_l) . X_l files the values of closed questions common to all samples. Y_l files the frequency of each word in each respondent answer. The issue arising from vocabularies which differ according to languages is solved through integrating the common closed questions:

- Y_l are separately transformed into proportion matrices P_l and columns margins are filed into a diagonal matrix D_l .
- Tables X_l are centred for the row weights issued from the row margins of Y_l and juxtaposed column-wise into matrix X . The column-variables of X are summarized through their principal

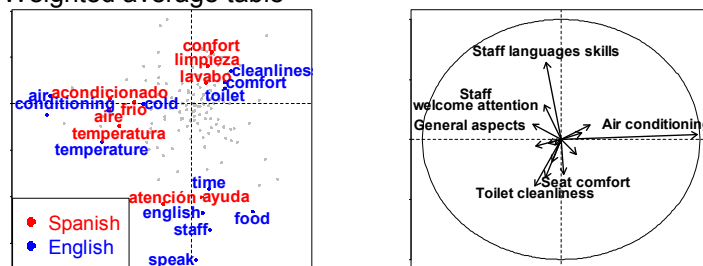
components filed in matrix Ψ . Δ , diagonal matrix with the eigenvalues, is the correlation matrix of the Ψ columns. Ψ can be considered as the juxtaposition of the Ψ_l submatrices corresponding to the individuals that have answered in language l ($l=1, \dots, L$).

- A dual multiple factor analysis (DMFA; Lê and Pagès 2010) but adapted to weighted average tables is applied to the global matrix juxtaposing column-wise the matrices $D_l^{-1}P_l'\Psi_l\Delta^{-1}$ with metrics D (diagonal matrix filing all the D_l) and Δ . Note that $D_l^{-1}P_l'\Psi_l\Delta^{-1}$ places the words on the principal components by weighted averaging, after sphericising the cloud of individuals by postmultiplicating Ψ_l by Δ^{-1} .

A global representation of words and variables on principal axes is provided.

Two samples of English and Spanish-speaking train passengers have scored travel characteristics and mentioned aspects to be improved through free comments. In the figure hereafter, the first axis underlines the dissatisfaction of the passengers for *air conditioning* as they strongly insist on it in their free comments. The second axis opposes problems linked to *staff* (general problems for both samples; insistence on the lack of English "speaking" staff by English speakers) and to cleanliness and comfort.

Keywords: Dual multiple factor analysis, Open-ended questions, Weighted average table



References

S. Lê and J. Pagès. DMFA: Dual Multiple Factor Analysis, *Communications in Statistics - Theory and Methods*, 39, 3, 483-492, 2010.

General efficiency modeling with planned preventive maintenance

Makram KRIT

Université de Gafsa, Institut Supérieur d'Administration des Entreprises de Gafsa, Tunisia

The aim of this paper is to propose a general model of the joint effect of corrective and preventive maintenance on repairable systems. The failure process without maintenance is characterized by intensity in

bathtub form. This form presented as a superposition of two Non-Homogeneous Poisson Processes (NHPP) and Homogeneous Poisson one (HPP). The maintenance effect is expressed by the change induced on the failure intensity before and after maintenance. The modeling frameworks are proposed and studied in order to generalize several models to the planned preventive maintenances. The generalization of the models suggested makes it possible to integrate the dependence between corrective and preventive maintenances. The likelihood functions are derived, so parameter estimation and assessment of the maintenance efficiency are possible. The properties of the parameters estimators have to be theoretically studied. Finally, results are applied to a real maintenance data set.

Keywords: reliability, maintenance, reparable system, stochastic modeling, failure intensity, estimation, likelihood.

BPSO Versions with Chi-squared Distribution for MKP Resolution

Raida KTARI^{1,2}, Habib CHABCHOUB¹

¹Sfax University, LOGIQ, Tunisia

²Aix-Marseille Université, CNRS LIF UMR 7279, France

This paper deals with the application of BPSO (Binary Particle Swarm Optimization), EPSO (Essential Particle Swarm Optimization) and EPSOq (Essential Particle Swarm Optimization queen) to the Multidimensional Knapsack Problem (MKP) which is well-known to be NP-hard Combinatorial Optimization problem. The particularity of this paper consists in proposing a novel random number generator based on the Chi-squared distribution for the particles' positions and velocities in the initialization step. A repair operator is also utilized to change an unfeasible solution to a feasible one. The performance assessment of the BPSO, EPSO and EPSOq is a critical point in this paper. That is why, we experiment these approaches on a variety of MKP instances from OR-Library. Then, we compare our results with those of other previous works and with the best known results in the literature.

Keywords: MKP, PSO, Chi-squared distribution, randomness, repair operator, performance.

Generalized Asymmetric Linnik distributions and process

Kuttykrishnan A.P.

Kannur University, India

In recent years there has been an increasing interest in developing the theory and applications of geometric stable distributions. The class of geometric stable distributions is a four-parameter family of distributions denoted by $GS_{\alpha}(\sigma, \beta, \mu)$ and conveniently described in terms of characteristic function

$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha} \varpi_{\alpha, \beta}(t) - i\mu t} \quad \text{where}$$

$$\varpi_{\alpha, \beta}(t) = \begin{cases} 1 - i\beta \text{sign}(t) \tan(\pi\alpha/2) & \text{if } \alpha \neq 1 \\ 1 + i\beta \frac{2}{\pi} \text{sign}(t) \log|t| & \text{if } \alpha = 1. \end{cases}$$

The parameter $\alpha \in (0, 2]$ is the index of stability and determines the tail of the distribution. These classes of distributions arise as a limiting distribution of geometric random sums of independent and identically distributed random variables. Since the geometric random sums frequently appear in many applied problems in various areas, the geometric stable distributions have wide variety of applications especially in the field of reliability, biology, economics, financial mathematics etc. When $\beta = 0, \mu = 0$, the geometric stable distribution have the

$$\Phi(t) = \frac{1}{1 + \sigma^{\alpha} |t|^{\alpha}},$$

characteristic function and the corresponding distribution is called Linnik distribution and is named after Ju.V.Linnik, who showed that the above function is a bona fide characteristic function of a symmetric distribution for any $0 < \alpha \leq 2$. It may be noted that probability density and distribution functions of the Linnik random variable are not in closed form except for $\alpha = 2$, which corresponds to the Laplace distribution. The Laplace distribution is symmetric, and there were several asymmetric extensions in generalizing the Laplace distribution.

In this paper we introduce and study new classes of distributions, namely Pakes generalized asymmetric Linnik distribution and geometric Pakes generalized asymmetric Linnik distribution. First order autoregressive process with Geometric Pakes generalized asymmetric Linnik

distribution as marginal distribution is developed. Higher order extensions are discussed. A bivariate distribution related to geometric Pakes asymmetric Laplace and Linnik distribution is introduced and bivariate time series model corresponding to this distribution is developed.

Keywords: Autoregressive process, Geometric infinite divisibility, Geometric exponential distribution, Geometric marginal asymmetric Laplace and Linnik distribution, Geometric Pakes generalized asymmetric Linnik distribution, Geometric Stable distribution.

One dimensional embedding for nonnegative data visualization

Lazhar Labiod, Mohamed Nadif

Department of Computer sciences, LIPADE, University of Paris 5,
France

Data visualization has attracted more attention this last decade as a powerful tool for a better understanding of data. In this paper we propose a new theoretical framework for data visualization, this framework is based on Rank-one SVD looking up an appropriate leading left and right singular vectors of a rows or=and columns normalized m by n adjacency matrix A . This involved computing of a truncated rank one singular value decomposition of a suitable normalized data matrix, constructing an one dimensional embedding for both rows and columns data.

The visualization of A consists in a simple permutation of rows and columns data according the sorted first left and right singular vectors, which involves an optimal data reorganization revealing homogeneous blocks. Finally, we link our approach to spectral co-clustering and show its usefulness in the context of co-clustering.

Keywords: Data visualization, Stochastic data, Power method, SVD, Co-clustering.

Multi-objective micro genetic algorithms. Two new approximations

Lacruz Beatriz, Lahoz David, Mateo Pedro M.

Depto. Métodos Estadísticos, Universidad de Zaragoza, Spain
uG-ELM is a bi-objective micro genetic Extreme Learning Machine (ELM) algorithm developed by the authors for learning single-hidden layer feedforward neural networks (SLFN) which exhibits a good

performance. In this algorithm we minimized the number of hidden neurons and the Mean Square Error (MSE) simultaneously in order to get the best neural network. It has been compared with the following ELM's algorithms: Original ELM, EOS-ELM, OS-ELM, OP-ELM and EI-ELM.

In this work we propose the modification of our algorithm by means of the definition of two different multi-objective approximations. First we propose to incorporate the standard deviation of the MSE together with the MSE and the number of hidden neurons in order to get a three-objective micro genetic ELM which also takes into consideration the dispersion of the errors to rank the solutions. In a second tentative we present a similar objective function as the original one but perturbing the MSE with a controlled noise which depends on its standard deviation.

Our aims with these modifications is reaching SLFN with a higher generalization capacity, for testing that, the results of the new algorithms will be compared with the obtained by means of the original uG-ELM.

Keywords: Multi-objective Optimization, Artificial Neural Networks, Genetic Algorithms

Modelling death rates with stochastic differential equations: an innovative approach

Sandra Lagarto, Carlos A. Braumann, Dulce Gomes

Universidade de Évora / CIMA – Colégio Luís Verney, Portugal

To explain mortality trends in human population and to predict its behavior in the medium / long term, we consider a cross-sectional analysis over time, instead of doing a cohort data analysis, and assume that there are random environmental fluctuations that may explain the variability of the phenomenon.

Considering these facts, we apply stochastic differential equations models (with a long tradition of application on finance and, more recently, on animal growth models) to crude death rates (CDR) of Portuguese population (annual time series available from 1940 to 2009 for selected ages 0 to 99, by sex).

Additionally when analyzing data across time, we could see that death rates had a strong decreasing trend during the last century which suggested the use of very simple one-dimensional models, like a Geometric Brownian Motion Model or a Generalize Stochastic Gompertz Model, applied to a certain age (selected ages from 0 to 99) and sex. Also we noticed that there were similar patterns in the data (by age and sex) which suggests the application of multivariate models with a correlation structure.

We show, in particular, an application of a bi-dimensional Stochastic Gompertz model (BSGM) with correlated Wiener processes by sex, for each considered age [1]. To compare results we use a first order vector autoregressive model, VAR(1).

Keywords: Mortality rates, cross-sectional analysis, multivariate SDE models, VARMA models.

[1] Lagarto, S., Braumann, C. (2012). Modeling human population death rates: A bi-dimensional stochastic Gompertz model with correlated Wiener processes. Studies in Theoretical and Applied Statistics. Springer Series (in print)

Kernel Regression Model with Correlated Errors

Dagmar Lajdová, Jan Koláček, Ivanka Horová

Department of Mathematics and Statistics, Masaryk University, Czech Republic

Kernel methods represent one of the most effective nonparametric methods for the estimation of a regression function. But there is a serious difficulty connected with them - the choice of a smoothing parameter called a bandwidth. Standard methods for selecting the bandwidth of a kernel estimator from data perform badly when the errors are correlated. There exist several possibilities for overcoming this. The partitioned cross-validation method and the plug-in method are presented. We compare these methods on both simulated and real data.

Keywords: Kernel, regression, bandwidth selection, correlated errors

Approach for Vehicle Routing Problem with Resource Constraints

A. LAMAMRI¹, A. NAGIH², H. AIT HADDADENE²

¹UKM, Algeria, ²Laboratory LITA, University Paul Verlaine Metz, France

The vehicle routing problem with resource constraints is a generalization of the classical vehicle routing problem. Given the large size of the problems encountered in practice, these models are solved by an approach based on column generation that can handle implicitly all feasible solutions and a master problem determining the best solution. In this paper, we present the development of a new approach to improve the acceleration of the method of column generation for solving the problem of construction vehicle routing, it is projected in each arc, the resources a vector of size smaller by using a Lagrangean relaxation algorithm to determine the coefficients of the projection arc combined

with an algorithm for re-optimization, then generates a sub-set of solutions to the master problem. Numerical testes on problems from instances of random vehicle routing giving competitive results.

Keywords: Combinatorial Optimization, Vehicle routing problem, path problems in graphs.

Forecasting admissions of Comsats Institute of Information Technology (CIIT)

Noor.M.Larik

COMSATS, Pakistan

The increasing competition for admissions in Higher Education Institutes has become a problem for Higher Education Institutes for making adequate policies for decision making. The COMSATS Institute of Information Technology has eight campuses all over the country and headquarters in Islamabad, Pakistan. Due to its popularity and ranking first in the country as institute of science and technology and sixth in all general universities of country, there has been a significant increase of the students for admissions. Thus, it requires forecasting of future demand for admissions to make adequate policies for students as well as faculty requirement.

In this article, an attempt has been made to forecast admissions of students particularly in main campus located in Islamabad by using models such as Holt's Linear Trend Model, Simple Linear Regression Models and Simple Linear Model. From the results it appears that Holt's linear trend model gives more appropriate forecasting results as compared to others.

Ruin probabilities for risk models with ordered claim arrivals

Claude Lefèvre¹, Philippe Picard²

¹Université Libre de Bruxelles, Département de Mathématique, Belgium

²Université de Lyon 1, ISFA, France

Recently, Lefèvre and Picard (IME 49, 512-519 (2011)) revisited a non-standard risk model defined on a fixed time interval $[0, t]$. The key assumption is that, if n claims occur during $[0, t]$, their arrival times are distributed as the order statistics of n i.i.d. random variables with distribution function $F(s|t)$ for $0 < s < t$. The present paper is concerned with two particular cases of that model, namely when $F(s|t)$ is of linear form (as for a (mixed) Poisson process), or of exponential form (as for a linear birth process with immigration or a linear death-counting process). Our

main purpose is to obtain an explicit formula for the non-ruin probabilities in these cases. This is done by exploiting properties of an underlying family of Appell polynomials. The ultimate non-ruin probabilities are then derived as a limit.

Keywords: Order statistics; (mixed) Poisson process; linear birth or death process; non-ruin probability; finite or infinite horizon; Appell polynomials

Stochastic Gene Expression in Prokaryotes: A Point Process Approach

Emanuele Leoncini^{1,2}

¹INRA - MIG, ²INRIA Paris, France

The gene expression is a fundamental process in cells and crucial for life. The inherent stochasticity of gene expression lead to first stochastic characterisations since the late 70s, but only in the last few years reliable experiments have given first quantitative results of proteins numbers and variances in single cells.

The stochastic uctuations of the number of copies of a given protein are investigated and a stochastic model representing the main steps of the production of a protein is presented: gene activation, transcription and translation. In particular, we revisit and generalize the current mathematical results investigating the properties of intrinsic noise in gene expression and, in the biological context of our interest, the theoretical results have allowed to obtain closed form expressions of the mean and variances of all the processes as a function of the model parameters.

Gene activation. The gene activation involves complex processes and is usually described as a telegraph process. In particular, the gene activation and inactivation are an exponentially distributed processes.

Transcription. When the gene is active, it produces a mRNA at rate λ_2 . A mRNA has a lifetime which follows now a general distribution $F_2(dy)$.

Translation. A given mRNA produces proteins at rate λ_3 and $F_3(dy)$ is the general distribution of the lifetime of a protein.

The classic modeling strategy is to consider a Markovian description of the protein production, which requires each step to be exponentially distributed. In this talk a more realistic model is considered. In this setting a simple Markovian description is not available and the classic PDE approach, based on Fokker-Plank equations, can no longer be used. Using an approach based on marked Poisson point processes, general results on the distribution of the number of proteins can be derived. A marked Poisson point process can be described as a classical Poisson process where each point has a mark which is an element of some i.i.d. sequence of random variables. For example, in the case of

the proteins produced by a specific messenger, the proteins dynamics can be described by the sequence $(s_n; \sigma_{3,n})_{n \in \mathbb{N}}$, where the sequence of protein births (s_n) is a Poisson process with parameter λ_3 and $(\sigma_{3,n})$ is the sequence of the protein lifetimes with a general distribution $F_3(dy)$. Up to present, only the case in which F_3 is exponentially distributed has been investigated. In practice, the distribution F_3 of a protein lifetime is more likely to be concentrated around a small neighbourhood of some constant. Our method, by using an analytic formula of the generalized Laplace transform of a marked Poisson point process, gives the first moments of the number of proteins at equilibrium in a general context, in particular when where the protein lifetime has a general distribution. Analytic closed form expressions are obtained for the mean number and the variance of the number of protein. When the mean lifetime is fixed, our results show in particular that, surprisingly, the variance of the number of proteins results increased with respect to the exponential assumption. This counter-intuitive result shows that the classic assumption of exponentially distributed mRNA and protein lifetimes leads to underestimate the protein variance. Various stochastic assumptions are discussed as well as the impact of the dilution process in the cell. The MPPP approach turns out to be a well adapted tool to describe gene expression, allowing a more realistic description of the phenomena still giving analytic formula of the first moments of the number of proteins.

Keywords: Stochastic gene expression, Poisson point processes.

Joint work with: Vincent Fromion (MIG team - INRA - vincent.fromion@jouy.inra.fr) and Philippe Robert (RAP team - INRIA - philippe.robert@inria.fr).

Adult Mortality Estimates for Small Areas in Brazil, 1980-2010

Everton Lima; Bernardo Queiroz; Flavio Freire; Marcos Gonzaga

Dept of Demography, Universidade Federal de Minas Gerais, Brazil

Background

In recent decades, life expectancy and health status have improved considerably in Brazil, life expectancy at birth has risen from 48 years in 1960 to over 73 in 2010. However, the country is marked by large regional and socioeconomic inequality. Due to data limitations; incomplete coverage of vital registration systems, errors in age declaration for both population and death counts; little is known about the level and trends in adult mortality for sub-national areas. This paper aims to study the evolution of adult mortality in small areas in Brazil from 1980 to 2010. We use 45q15 as a summary measure of adult mortality across small areas in Brazil.

Data and Methods

We propose a methodological approach that combines the death distribution methods (DDM) to indirect standardization and Bayesian statistics to produce more reliable estimates of adult mortality for small areas. We make use of the mortality database available from Datasus. Population by age and sex comes from national population censuses. The data are organized for 137 comparable small areas from 1980 to 2010.

Finds and interpretations

The results that data quality has improved considerably across the country during the period of analysis, but regional variation persists. On average, completeness of death counts increased from 80%, on average, in 1980-1991 to 95%, on average, in 2000-2010. Estimates of adult mortality shows that more developed regions in the South and Southeast parts of the country have lower mortality levels than the Northeast, but there is a clear convergence process.

Conclusions

The paper indicates that the combination of traditional demographic methods to Bayesian statistics is a good alternative to produce mortality estimates in small areas in Brazil. We also find a rapid improvement in the quality of mortality data in Brazil since 1980. Finally, we observe a convergence process on health status.

Keywords: small areas, adult mortality, mortality estimation, Brazil

Sample coefficient of variation under uncertainty

Shiang-Tai Liu

Graduate School of Business & Management, Vanung University,
Taiwan

The coefficient of variation is a dimensionless number that quantifies the degree of variability relative to the mean. The coefficient of variation is a useful statistical measure that has long been widely used as a descriptive and inferential quantity in many applications of science, economics, and other areas. In real world applications, the observations in the model might be inexact and imprecise in nature. There are also situations that measurements are never 100% accurate, there is always a measurement inaccuracy. One approach for dealing with uncertainty in observations to represent uncertain observations in real world applications is by intervals. These studies reveal a fact that when a sample with numbers of uncertain observations, which are presented as interval values, the calculated results should be an interval value as well; that is, lying in a range. Different from the previous studies, which design! the heuristic algorithms to calculate the possible values of mean and variance, we construct a pair of mathematical programs to calculate the lower bound and upper bound of the sample coefficient of variation

with interval-valued observations. By variable substitutions, the mathematical programs are transformed into a pair of quadratic programs. Solving the pair of quadratic programs generates the interval of the sample coefficient of variation. The obtained results may provide the decision makers with more information for making decisions. An example is illustrated to explain the whole idea proposed in this paper.

SGR modeling of fake ordinal data with correlational structures

Luigi Lombardi¹, Massimiliano Pastore², Massimo Nucci³, Andrea Bobbio⁴

¹Dipartimento di Psicologia e Scienze Cognitive, Università di Trento,

²Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, ³Dipartimento di Psicologia Generale, Università di Padova, ⁴Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata, Università di Padova, Italy

In many psychological inventories (i.e., personnel selection surveys and diagnostic tests) the collected samples often include fraudulent records. This confronts the researcher with the crucial problem of biases yielded by the usage of standard statistical models. In this paper we generalize a recent probabilistic perturbation procedure, called SGR - Sample Generation by Replacements - (Lombardi & Pastore [4]), to simulate fake data with correlational structures. To mimic these more complex faking data we proposed a novel extension of the SGR conditional replacement distribution which is based on a discrete version of the truncated multivariate normal distribution. We also applied the new procedure to real behavioral data on the role of perceived affective self-efficacy in social contexts.

Keywords: Sample Generation by Replacement, Fake-good data, Truncated multivariate normal distribution.

Some procedures for solving common difficulties in Bayesian Space-Time models

Alberto López, M. Pilar Muñoz, Josep A. Sánchez

Dept. of Statistics and Op. Research, Universitat Politècnica de Catalunya, Spain

In the field of Bayesian hierarchical space-time models, lots of papers are written for estimating and predicting the significant variables, as for example climate temperature, wind speed, etc. However, there are not

many articles related to the initial and intermediate steps of the process for estimating and predicting those variables. For example, how to define the theoretical model, the selection of the initial values, etc.

The objective of this work is to set up some norms or guides for helping and solving the most common difficulties of fitting space-time data to this methodology.

Specifically, this paper addresses: the correct theoretical model definition; the explanation of different spatial structures and how they affect the estimation algorithm; and, most importantly, provide alternatives for selecting the correct initial values for the hyperpriors parameters.

Finally, these procedures have been applied in two real data sets: temperature and wind speed, updated every three hours in Tarifa (Spain). The sensitivity analysis, which was introduced previously, has provided very accurate estimators from the large number of parameters that must be estimated in this class of models.

Keywords: Bayesian hierarchical space-time models, initial values, hyperpriors

Towards Dynamic Bayes Networks to model Time Series from the WatSan4Dev subset

**Jorge López Puga¹, Céline Dondeynaz^{2,3}, and César Carmona
Moreno²**

¹ Department of psychology, Catholic University San Antonio, Murcia,
Spain

² Institute for Environment and Sustainability, Joint Research Centre,
Ispra, Italy

³ School of Engineering, Liverpool University, Liverpool, United Kingdom

Water is the essence of life, without this biochemical molecule, life would not be possible on Earth. Water assumes a similar key role if considering human/society development: water supply heavy burden and lack of

sanitation slows down the economy, social improvement and population well-being. However, millions of people lack access to drinkable water (894 million) and/or basic sanitation (2.5 billion)¹.

A recent research originated in the Joint Research Centre – European Commission, has developed a coherent database (WatSan4Dev) to identify the most influential variables on Water Supply and Sanitation (WSS) services in developing countries. The WatSan4Dev dataset is constituted of official indicators provided by international agencies. The 25 variables selected are organized into five thematic areas (water resources, human activity pressure on water resources, country environmental concern, human development and official development aid) for 101 developing countries.

The objective of this work is to develop Dynamic Bayesian Networks (DBN) to model time series from the WatSan4Dev subset. Three temporal points (2000, 2004 and 2007) are used to create two different BN (one for water supply and another for sanitation) based on six composite variables. The BN set links between the three time slices to create a unique network. The error rates of the query variables (WSS) were below 11%, allowing using the networks to run and analyze different scenarios. This tool is useful to orient and support decision/strategy making related to WSS and beyond, to human development. Methods, results of the analysis but also progresses made towards dynamic models (DBN) will be presented.

Keywords: Water Supply and Sanitation services, Bayesian networks, time series, sustainable development, Millennium Development Goals.

.

¹ http://www.unwater.org/statistics_san.html

Adaptive Monitoring Rules based on Scan Statistics for Compound Patterns with an Application to Quality Control in Blood Banking

Wendy Lou¹, James C. Fu², Qilong Yi³

¹Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, ²Department of Statistics, University of Manitoba, ³National Epidemiology and Surveillance, Canadian Blood Services, Canada

Scan Statistics have been used widely in various biomedical applications, such as genomic sequence analysis, and also in public health sciences, including healthcare surveillance. This study is motivated by a practical application involving quality control of allogeneic blood components, where a set of fixed sampling criteria is used to test all blood production sites regardless of the number of blood products collected at individual sites. It is commonly found in real data that the variation of production ranges 5 to 15 fold across blood centers over the monitoring duration, and hence, for example, testing a fixed percentage of blood units will likely result in sub-optimal usable blood production. We propose a scan-like approach using compound patterns to provide a flexible adaptive strategy for determining the quantity of samples to be tested. The approach will allow varying the sampling criteria based on sequential monitoring, to minimize the number of testing samples and therefore optimize the usable total production of blood units. The proposed approach utilizes the concept of multiple waiting times, in the sense of runs and patterns, and the decision rules to modify the testing criteria can be considered under a Markov-dependent process. An example using real data from the quality testing of blood products will be presented to illustrate the methodology.

Keywords: Runs and Patterns, Sampling Inspection, Markov Chain.

Robust prediction of functionals of stochastic sequences with stationary increments

Maksym Luz, Mykhailo Moklyachuk

Department of Probability Theory, Statistics, and Actuarial Mathematics,
Kyiv National Taras Shevchenko University, Ukraine

The problem of optimal estimation of the functional $A\xi = \sum_{k=0}^{\infty} a(k)\xi(k)$

depending on the unknown values of a stochastic sequence $\xi(k)$ with stationary increments of order n [1] from observations of the sequence for the moments of time $k = -1, -2, \dots$ is considered. Formulas for calculation the spectral characteristic and the mean square error of the optimal linear estimation [2] of the functional $A\xi$ are proposed when the spectral density $f(\lambda)$ of the sequence is known and admits the canonical factorization

$$|\Phi_{\mu}(e^{-i\lambda})|^2 = \left| \sum_{k=0}^{\infty} \varphi_{\mu}(k) e^{-i\lambda k} \right|^2 = \frac{|1 - e^{-i\lambda\mu}|^{2n}}{\lambda^{2n}} f(\lambda).$$

In this case $h_{\mu}^{(a)}(\lambda) = A(e^{i\lambda}) - r_{\mu}^{(a)}(e^{i\lambda})\Phi_{\mu}^{-1}(e^{-i\lambda})$, where

$$A(e^{i\lambda}) = \sum_{k=0}^{\infty} (D^{\mu}a)_k e^{i\lambda k}, \quad r_{\mu}^{(a)}(e^{i\lambda}) = \sum_{j=0}^{\infty} (D^{\mu}A\varphi_{\mu})_k e^{i\lambda j},$$

$$\Delta(f, \bar{A}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |r_{\mu}^{(a)}(e^{i\lambda})|^2 d\lambda = \|D^{\mu}A\varphi_{\mu}\|^2.$$

Matrix A is defined by coefficients $a(k), k \geq 0$, $\varphi_{\mu} = (\varphi_{\mu}(0), \varphi_{\mu}(1), \varphi_{\mu}(2), \dots)$.

In the case where the spectral density $f(\lambda)$ is not known exactly, but, instead, a class of admissible spectral densities is given formulas that determine the least favorable spectral densities and the minimax (robust) spectral characteristics [3] of optimal estimate of the functional are proposed.

References

1. A.M.Yaglom. *Correlation theory of stationary and related random processes with stationary n th increments*, Math. Collection 37(79)(1955), no. 1, 141-196.
2. A.N. Kolmogorov. *Selected works of A.N. Kolmogorov. Vol. II: Probability theory and mathematical statistics*, Ed. by A.N. Shirayev. Mathematics and Its Applications. Soviet Series. 26. Dordrecht etc.: Kluwer Academic Publishers, 1992.

3. M.P. Moklyachuk, A.Yu. Masyutka, *Minimax-robust estimation technique. For stationary stochastic processes*, LAP LAMBERT Academic Publishing, 2012.

Asymptotic results for empirical means of independent geometric distributed random variables

Claudio Macci, Barbara Pacchiarotti

Dipartimento di Matematica, Università di Roma Tor Vergata, Italia

The theory of large deviations gives an asymptotic computation of small probabilities on exponential scale. Estimates based on large deviations play a crucial role in resolving a variety of questions in statistics, engineering, statistical mechanics, applied probability and simulation.

The aim of this paper is to prove large and moderate deviation results for empirical means of independent geometric distributed random variables. The main tool is Gartner Ellis Theorem and in a case the proof is quite intricate because we have some non-exposed points. We also present applications of the large deviation results for the weak record values of i.i.d. discrete random variables.

Keywords: large deviations, moderate deviations, Gartner Ellis Theorem, exposed points, weak records.

Emulation of Complex Simulator Models with Application to Hydrology

David Machac, Carlo Albert, Peter Reichert

EAWAG, Switzerland

Emulation is a technique, where a deterministic simulator is replaced by its stochastic approximation, an emulator. The aim of the emulation is then to reduce the evaluation times of a simulator. This proves useful in cases, where we need to do a large number of simulator runs up to the point where it becomes infeasible. For this technique, the simulator has to be evaluated only for certain input parameter sets to generate outputs used for the conditioning of a Gaussian process prior, which is the basis of an emulator. The conditioning has to be done only once and the number of input parameter sets used for it depends on the required accuracy of the emulation. In this work, we reduce the number of these conditioning evaluations while keeping the accuracy by using our knowledge of the underlying mechanisms of the simulator. This work focuses on emulating an urban drainage simulator Storm Water

Management Model, which uses the shallow water equations to simulate the water routing.

Keywords: Emulation, stochastic approximation, Gaussian process

A bayesian approach for PARAMETER ESTIMATION OF A plant growth model

Sonia Malefaki¹, Trevezas Samis², Paul-Henry-Cournède²

¹Department of Engineering Science, University of Patras, Greece

²EPI DIGIPLANTE, INRIA, Laboratory of Applied Mathematics and Systems, Ecole Centrale Paris, France

Plant development and growth involve complex mechanisms. The parameterization and parameter estimation of models describing these phenomena are very challenging problems. The last decades advanced plant growth models have been proposed in the literature and among them the GreenLab model (Cournède *et al.*[1]) has been proved to be very promising. This study focuses on parameter estimation issues for individual plant growth models based on the GreenLab formulation. The available data consist of organ masses measured only once by censoring plant's evolution at a given observation time (destructive measurements). An interesting approach to take into account modeling errors was proposed in Trevezas and Cournède[2] which describes data sequentially and the evolution of a certain class of plants with known structural development as non-homogeneous hidden Markov models. The first parameter estimation approaches were based on stochastic variants of the EM algorithm (Trevezas, Malefaki and Cournède [3]). In this study, we present a Bayesian approach for estimating the model's parameters. The Bayesian approach has much more flexibility in handling complex structures providing a useful tool for analyzing this type of models. Moreover, it allows the use of prior information in addition to the information included in the observed data for producing more accurate results. The Bayesian estimation is particularly recommended for small samples (frequently encountered in this application) in contrast to maximum likelihood estimation (MLE) which has gained its popularity due to its asymptotic properties. An MCMC approach is adopted to approximate the posterior distribution of the parameters and several implementation issues are discussed. The performance of this method is illustrated in a real dataset from the sugar-beet and a comparison is made with the MLE approach.

Keywords: plant growth model, Bayesian approach, MCMC methods, sugar-beet.

References

1. Cournède P.-H., Kang M. Z., Mathieu A., Barczi, J.F., Yan, H.P., Hu, B.G. and De Reffye, P. Structural Factorization of Plants to Compute

their Functional and Architectural Growth. *Simulation* **82**(7), 427-438, 2006

2. Trevezas, S. and Cournède P.H. A sequential Monte Carlo approach for MLE in a plant growth model, (*submitted*).

3. Trevezas S., Malefaki S. and Cournède P. H. Simulation techniques for parameter estimation via a stochastic ECM algorithm with applications to plant growth modeling, (*submitted*).

MMR & IMR in Raichur District: A Study Among Rural Poor

Prakash Malin, B M Ramesh

KHPT, District Health & Family Welfare office, India

This paper aims to study Mortality of Infant & Mothers in Rural Raichur district. The data is based on HMIS & District report & review report or investigation reports.

The major findings of this paper is that most of the deaths are happening in delay in check up or referring to the higher facilities. More over age of the women is less than 25. The cause of death for mothers are PPH, not completion of 3 ANC Checkups more over delay in treatment & referral.

Keywords: Raichur, Mortality, HMIS, IMR, MMR, PPH ect

The use of MDS and HCA enabled pharmacists to reveal their roles which reflect on country of practice and cultural differences when improving patients' adherence to asthma medication

Andrea Manfrin¹, Cecilio Mar Molinero², Luca Cacciolatti²

¹Medway School of Pharmacy, University of Kent, ²Kent Business School, University of Kent, UK

Pharmacists are health care professionals who have a lot more to offer to the public rather than just selling pills. Pharmacists live a dichotomy role: health care consultants versus shop keepers. The aim of this research was to assess how pharmacists intervene improving patient adherence to asthma medication and to see how pharmacists perceived themselves. The research was conducted by means of a multiple choice questionnaire that was completed by 551 pharmacists. The replies were analysed using Ordinal Multidimensional Scaling and Cluster Analysis, with the routine PROXSCAL. Pharmacists revealed their roles in terms of four dimensions: a community health orientation, theoretical-applied drug orientation, the degree of involvement with the patient and the role of the

pharmacist as a consultant. These dimensions have different salience in each one of the countries.

Keywords: pharmacy, country, asthma, adherence multidimensional scaling.

Typical Distances in Large Synchronized Stochastic Particle Systems

Anatoly Manita

Department of Probability, Faculty of Mathematics and Mechanics,
Lomonosov Moscow State University, Russia

We consider the interacting particle system $x(t) = (x_1(t), \dots, x_N(t))$, $t \in \mathbf{R}_+$, consisting of N identical Brownian particles with stochastic synchronization-like interaction. We provide the answer to the question: what is a typical distance between particles in the synchronized system (when $t \rightarrow +\infty$).

It appears that in some sense a distance between any pair of particles x_i and x_j is of order N .

We obtain some asymptotical results about exponential moments of the distances and show that distributions of $(x_i - x_j) / N$ have exponential tails uniformly in N .

Keywords: Stochastic synchronization, Brownian particle system, Markov process, renewal process.

Vibrations control: singular solutions for mean square optimization

Larisa Manita

National Research University Higher School of Economics,
Moscow Institute of Electronics and Mathematics, Russia

We consider a control problem for longitudinal vibrations of a nonhomogeneous bar with clamped ends. The vibrations of the bar are controlled by an external force which is distributed along the bar. For the minimization problem of mean square deviation of the bar we construct optimal solutions in the form of the Fourier series. To find Fourier coefficients we consider an optimal control problem in the space l_2 . For the control problem in l_2 we show that in a certain neighborhood of the origin the structure of the optimal solutions is the following one: for the finite time the optimal nonsingular trajectory enters the singular surface with infinite numbers of control switchings, after that the optimal trajectory remains on the singular surface and attains the origin for the infinite time.

Multivariate methods in historical data: An Optimal Scaling Analysis with Missing Categorical Data

Cecilio Mar Molinero¹, Fabiola Portillo²

¹University of Kent, UK, and Universitat Autònoma de Barcelona, Spain,

²Department of Economics and Business, University of La Rioja, Spain

The information contained in a database from the Spanish railway company MZA for the period 1882-1889 is explored by means of Optimal Scaling methods. The sample includes 992 employees who joined the Madrid-Atocha workshop during that period and contains missing qualitative values. A technique that combines correspondence analysis with the k-means clustering algorithm is implemented to impute these values, maximizing internal consistency as measured by Guttman's squared correlation ratio. The results show two characteristics observed in other studies of labour relations: the existence of «ports of entry» for workers at low levels of qualification and long-term labour relations.

Keywords: Optimal Scaling Analysis; Missing Categorical Data; Labour markets.

Admissible strategies and European claims for non-Markovian SDEs

Ionela Marinela Marinescu

Department of Applied Mathematics, The Bucharest University of Economic Studies, Romania

In this paper we are seeking representations for the conditional mean

$E(\varphi(x(T)) | y^T)$ under the incomplete knowledge of the state $x(t), t \in [0, T]$ of a nonlinear stochastic differential system when

$\varphi(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a locally Lipschitz continuous function. Using a new probability measure, we define a new stochastic dynamical system. The solution of this new system is a continuous and weak differentiable function. Finally, we can define an admissible strategy in feedback form associated with financial market. The results obtained are then used to define an admissible strategy for a non-Markovian SDE. The solution of

this equation can be associated with the value function $V_\theta(t)$, $t \in [0, T]$, written for an admissible strategy in a financial market.

Keywords: non-Markovian SDE, admissible strategies

Acknowledgement: This research was supported by CNCS-UEFISCDI, Project number IDEI 303, code PN-II-ID-PCE-2011-30593.

Relationships between stock and bond markets in the Euro Zone, during crisis periods

M. Dolores Márquez, M. Pilar Muñoz, Pilar Soriano

Dep. Economics and Economic History, Universitat Autònoma de Barcelona, Spain

The aim of this study is to investigate the relationships between stock and bond markets in the Euro zone and how crisis periods affect the volatility of both markets. Then, we propose a regression with a Markov switching model to explain how the economic environment affects the evolution of the dynamic correlations.

To analyse the relationships between the markets and the volatility associated to the crisis periods, we estimate the dynamic conditional correlation with a Multivariate GARCH model, using two proxy indicators of the markets: A portfolio formed from the S & P Eurozone Government Bond with different maturities and the S & P Euro Stock Index.

Once the dynamic correlations have been estimated, we examine the macroeconomic variables, which are related to economic uncertainty, wealth and monetary policy. The introduction of Markov switching in the regression model improves the goodness of fit and identifies two regimes which are distinguished between crisis periods and non-crisis periods. Furthermore the estimated model explains how and which macroeconomic variables affect the evolution of the correlation.

Keywords: dynamic conditional correlation, markov switching, sovereign bonds, stocks returns, crisis

The Coxian Phase type distribution as a contribution in the multilevel model

Adele H. Marshall¹, and Mariangela Zenga²

¹ Centre for Statistical Science and Operational Research (CenSSOR)
Queen's University of Belfast
Belfast, BT7 1NNA, UK

² Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via B. degli Arcimboldi, 8,
20126 Milano, Italy

The Coxian Phase Type distribution (CPT) describes a latent Markov chain in continue time with one absorbing state and k latent transient

states. The time of the Markov processes may be divided into intervals according to the estimated phases (Marshall and McClean, 2003). Such intervals could be thought for discrete or grouped survival model and readily used into multilevel models where there are multiple failure times per subject, then the level-2 units are the subjects and the level-1 units are the failure times.

This paper aims to investigate the possible use of the CPT to individuate the latent intervals that will be used in a grouped survival multilevel model.

Keywords: Coxian Phase type distribution, Grouped survival model, multilevel model.

Early Forecasting of Parliamentary Seats Distribution: the Representative Polling Stations Method

Ana Martín-Arroyuelos, Karmele Fernández-Aguirre, Juan Ignacio Modroño-Herrán

Departamento de Economía Aplicada III and BETS: Bilbao Economic Theory and Statistics (UFI 11/46), Facultad de Ciencias Económicas y Empresariales, Universidad del País Vasco (UPV/EHU), Spain

On election nights, once the polling stations are closed, there is an evident interest in early estimation of the number of parliamentary seats obtained by each political party, based on actual counting of votes. This paper focuses on the representative polling stations method to forecast the aforementioned number of seats. This method involves selecting a reduced group of stations whose aggregate result turns out to be a sufficiently accurate prediction of the final distribution of seats. This paper presents a method based on past results useful to select such representative polling stations and then able to choose the needed proportion of them that should be considered. Emphasis is set out on forecasting accurateness.

Keywords: Parliamentary seats prediction, Representative polling stations, Distances, Linear Regression

Couple-wise divorce rates in Japan

Shigeru Mase, Kouji Yoshimoto

Department of Mathematical and Computer Sciences, Tokyo Institute of
Technology, Japan

In Japan, ratios of numbers of divorce to those of marriage per year exceed one-third from 2000 onward. This fact is frequently referred to as if onethird of married couples will divorce in future. This interpretation is doubtful since numbers of marriage has been rapidly decreasing in Japan because of declining birth rates and, on the other hand, numbers of divorce include those of married couples from older generations with higher birth rates. What are true couple-wise divorce rates, that is, rates of couples married in one year who will divorce afterward, in Japan? Estimates of these rates are important because short marriage durations will frequently result in children with insufficient parental cares and financial supports. In order to estimate couple-wise divorce rates, we have to forecast numbers of divorce in future for couples married each year. This can be done using methods to construct life table. The result shows that actual couple-wise divorce rates will be over one-third for couples married from 1997 onward, contradictory to our first hypotheses that actual couple-wise divorce rates are much less than one-third. Moreover, it will be almost 40% for those married in 2002.

Keywords: couple-wise divorce rates, vital statistics, Japan.

Modelling of Cohort Mortality Patterns – New Approaches

Petr Mazouch¹, Klára Hulíková Tesárková²

¹Department of Economic Statistics, Faculty of Informatics and Statistics, University of Economics, ²Department of Demography and Geodemography, Faculty of Science, Charles University in Prague, Czech Republic

The mortality modeling is one step for better forecasting of the future development of mortality. Aim is to find relations and hidden regularities and patterns in the mortality development. There exist many various methods of mortality forecasting leading to more and less accurate outlooks. This paper introduces a simple alternative approach which is based on cohort mortality patterns modeling. Fundamental base of the proposed model is the assumption of constant change of the force of mortality between two following ages across cohorts. This assumption was verified on time series of Swedish data which are long enough for this purpose. The mentioned changes of mortality between ages were really nearly constant for all the cohorts and ages used for this verification and the created time series of mortality changes meet the requirements for further statistical modeling. Finding of the coefficients of mortality increase between two ages and its intrinsic patterns across generations could be used for its very simple application to not yet extinct generations and through that for the estimation of their future mortality development.

Keywords: Mortality modeling, Cohort mortality,

Phase-type Models for Costing Patient Care Provision in Integrated Health and Social Services

Sally McClean¹, Jennifer Gillespie¹, Lalit Garg², Maria Barton¹, Ken Fullerton³

¹School of Computing and Information Engineering, University of Ulster, UK, ²Computer Information Systems, University of Malta, Malta, ³School of Medicine, Dentistry, and Biomedical Sciences, Queen's University, UK

Stroke disease places a heavy burden on society, incurring long periods of hospital and social care, with associated costs and reduced patient and carer quality of life. In addition, stroke is a highly complex and heterogeneous disease with diverse outcomes and multiple strategies for

therapy and care, which are well suited to a modelling approach. Also patients typically move through various stages or phases of care leading naturally to phase-type modelling as a suitable representation of the care process. Previously we developed a modelling framework which classifies patients with respect to their length of stay (LOS) on the basis of known covariates such as gender, age, diagnosis and discharge destination; phase-type models can then be used to describe patient flows for each class. Also multiple outcomes, such as discharge to normal residence, nursing home, or death are possible.

We here focus on incorporating costs into the previous phase-type models of the integrated care system and obtain results for moments of total costs. In addition we provide results for moments of total costs in $(0, t]$, for an individual, a cohort arriving at time zero and when arrivals are Poisson. Based on stroke patient data from the Belfast City Hospital we use the overall modelling framework to obtain results. Such a mixture approach is an effective methodology for prediction of costs in Markov systems where groups of individuals follow heterogeneous pathways. It is thus a powerful methodology for determining the relationship between input covariates and outcome measures and their interrelations.

Keywords: Markov models, Phase type distributions, Patient Costing, Length of stay.

Data analysis and reliability models of pumping stations of the haoud Elhamra-Bejaia pipeline (Algeria)

¹*Medjoudj Rafik, Imine Aghiles², ¹Medjoudj Rabah, ²Aissani Djamil

¹Electrical Engineering Department, University of Bejaia, ²LAMOS Laboratory, University of Bejaia, Algeria

in this paper, a case study for reliability data analysis of pumping stations of a pipeline is introduced. By modeling the times to failure and the mean times between failures of components at these stations, the Weibull distributions were obtained on the basis of Kolmogorov Smirnov tests. When analyzing the shape parameters of the obtained distributions, it was observed that their values are less than unity, indicating that, theoretically, the considered components are in a youthful period or that the components have been operating for about 25 years. These statements are contradictory with practical considerations (e.g., an aging period) and the shape parameters values should be greater than the unity. Regarding also the multiple failure causes of the considered components, the adjustment of data using mixed Weibull distributions is deemed adequate by the tests and the shape parameters values are globally greater than the unity and the trend confirms the nature of aging equipment. The reliability evaluation of a station's components and the

reliability change for different variants of the pipeline have been ascertained.

Keyword: Pumping station, Reliability, Weibull distribution.

On the nonlinear dynamics origin of intermittency in turbulent edge plasma of tokamaks

A. Mekkaoui, Y. Xu

Institute for energy and climate research, Forschungszentrum Jülich,
Germany

Intermittent events, typically high density plasma filaments (blobs) are observed in edge plasma of magnetic fusion devices (tokamak, e.g., JET, DIII-D, ITER). Many advance in turbulence diagnostics (e.g. gas puff imaging and Langmuir probe) enable experimentalist a wealthy access to statistical properties of fluctuations. These properties are found to be very common to several machines. In particular the plasma density fluctuations are positively skewed and are well fitted by a gamma distribution. Our purpose is the analysis of plasma density fluctuations in the TEXTOR tokamak, using strong criteria instead of data fitting. This could help us to focus on the relevant distribution candidate (at least confirming the gamma pdf) and emphasize on the underlying Fokker-Planck and Langevin equations. A stochastic process is then deduced from these experimental time series, and used as a turbulence model in edge plasma! modeling (e.g. for the design of ITER and DEMO). The latter consists in coupling monte-carlo transport code for neutral particles and plasma turbulence code. These investigation open the door to another view of experimental time series in tokamak turbulence. It consist in the possibility of using realistic time series to write down a turbulence equation. Which is just the opposite of the approach followed up to now, in the sens that experimental time series have been often used to confirm theoretical turbulence models.

Keywords: Intermittent transport in edge plasma of tokamak

References:

D. A. D'Ippolito and al., Phys. Plasmas, 18, 060501 (2011)

A. Mekkaoui, Phys. Plasmas, 20, 010701 (2013)

Spatial generalised linear mixed models based on distances

Oscar O. Melo, Carlos E. Melo, Jorge Mateu

Department of Agronomy, National University of Colombia, Colombia

Risk models derived from environmental data have been widely shown to be effective in delineating geographical areas of risk because they are intuitively easy to understand. We present a new method based on distances, which allows the modelling of continuous and non-continuous

random variables through distance-based spatial generalised linear mixed models (DBSGLMM). The parameters are estimated using Markov chain Monte Carlo (MCMC) maximum likelihood, which is a feasible and a useful technique. The proposed method depends on a detrending step built from continuous or categorical explanatory variables, or a mixture among them, by using an appropriate Euclidean distance. The method is illustrated through the analysis of the variation in the prevalence of Loa loa among a sample of village residents in Cameroon, where the explanatory variables included elevation, together with maximum normalised-difference vegetation index (NDVI) and the standard deviation of NDVI calculated from repeated satellite scans over time.

Keywords: Distance-based methods, Epidemiological study, Markov chain Monte Carlo, Spatial generalised linear mixed models, Spatial interpolation

Longitudinal data analysis distance based using generalized estimating equations

Sandra E. Melo, Oscar O. Melo, Carles M. Cuadras

Department of Agronomy, National University of Colombia, Colombia

In this work, we propose a methodology to analyze longitudinal data through distances between pairs of observations (or individuals) with regard to the explanatory variables and distances between times used to fit not normally distributed response variables. We present the proposed model, and we develop the process of parameter estimation and hypothesis tests; this process is made using a mixture between generalized estimating equations (GEE) methodology and distance based method. Furthermore, the inference and the process of validating of the proposed model are performed. The information analyzed through the application is part of a conservation project, breeding and management of wild bees. These data were taken in a rural area of Acacias (Meta, Colombia). The objective in this application was studied the pollen foraging activity (understand foraging as the collection process, pollen, developed by bee species of *Melipona fasciata*) in five different nests, and thus, determined the effect that could have the temperature and relative humidity on the activity. The use of distance based with GEE strategy is useful to predict both missing data and future observations, which have lower errors of prediction than the traditional models.

Keywords: Gower distance, distance-based model, generalized estimating equations

Analysis of High Order Fuzzy Time Series by MLP

Memmedaga Memmedli, Ozer Ozdemir

Department of Statistics, Faculty of Science, Anadolu University,
Eskisehir, Turkey

Fuzzy set theory was introduced by Zadeh. Since then, many fuzzy time series models have been proposed such as first order models, high-order models, hybrid models and seasonal models. These fuzzy time series models have been applied to various problem domains, such as enrollment, temperature and stock index.

Artificial neural network(ANN) become effective tool for researchers by forecasting fuzzy time series. In forecasting, ANNs are mathematical models that imitate biological neural networks. There are various types of ANNs. One of them is Multilayer Perceptron (MLP).

MLP is constructed of multiple layers of computational units. Each neuron in one layer is directly connected to the neurons of the subsequent hidden layer. In many applications, the frequently used activation function is sigmoid function. MLP uses a variety of learning techniques, the most popular being back-propagation.

In this study we aimed to handle a nonlinear problem to apply neural network based high order fuzzy time series model. Differing from previous studies, we used a new method which used various degrees of membership in establishing fuzzy relationships with various numbers of hidden nodes and we performed MLP models to improve forecasting performance. To demonstrate comparison between these models we used Enrollment data of Alabama University which is the most commonly used in forecasting fuzzy time series and a large data set of exchange rate of Turkish Liras to Istanbul Stock Exchange National-100 Index for the years between 2010 and 2012 with comparing other methods in the literature.

Keywords: Forecasting, high order fuzzy time series, multilayer perceptron, artificial neural network, enrollment.

Assessing the Model Performance of Nonparametric Fuzzy Local Polynomial Regression with Different Bandwidth Selection Methods

Memmedaga Memmedli, Munevvere Yildiz

Department of Statistics, Faculty of Science, Anadolu University,
Turkey

Fuzzy regression studies mainly have focused on fuzzy parametric regression models. In practical problems, expressing relation between explanatory variable and response variable with a certain parametric

model is a very important restriction, and it can cause incorrect results. Hence, interest in nonparametric regression models have increased in recent years and for this purpose different type of models have been developed. For example k-nearest neighborhood regression, local polynomial models, kernel regression, different regression models with spline functions, etc. But, there are very few studies on fuzzy nonparametric regression models. It is necessary transform many notations and approaches of nonparametric regression models to fuzzy form for creating fuzzy nonparametric regression models.

In this study, we considered the relationship between the smoothing parameter value and degree of polynomial as a simulation study in nonparametric fuzzy local polynomial regression. Besides the local linear models, local cubic models are also used in this simulation study. Fuzzy forms of cross validation and generalized cross validation criteria are developed for bandwidth selection. Performances of the models are evaluated in accordance with the selected bandwidth. Simulation results showed the degree of relationship between the bandwidth and order of polynomial in experimental way.

As a result, it is obtained that the bandwidth size increases while order of polynomial increases. This leads to reduction of local fitting points and as a result overall operations decrease. Especially data with fluctuating; usage of local linear models are not convenient; in this case, it is necessary to reduce bandwidth but it increases computational complexity. Local bandwidth can increase while order of polynomial increases.

Keywords: Local polynomial smoothing, fuzzy nonparametric regression, bandwidth selection, cross validation, generalized cross validation.

A virtual age model for imperfect repair of a continuously monitored system with accumulating deterioration

Sophie Mercier¹, Inmaculada T. Castro²

¹Laboratoire de Mathématiques et de leurs Applications - Pau, IPRA, Université de Pau et des Pays de l'Adour, France, ²Department of Mathematics, Escuela Politécnica, University of Extremadura, Spain

A continuously monitored system is considered, which is subject to accumulating deterioration modeled as a gamma process. The system fails when its degradation level exceeds a limit threshold L . At failure, a signal is sent to a maintenance team which arrives after a delay τ . Once arrived, the maintenance team performs an instantaneous and perfect replacement. To shorten the system down time (with length τ), the following condition-based preventive maintenance strategy is applied: instead of waiting for the system failure, the signal is preventively sent to

the maintenance team at the reaching time of a preventive threshold M , (with $M < L$). If the system is not too deteriorated at the maintenance time, an imperfect repair is instantaneously performed. Mimicking virtual age models used for recurrent events (see [1]), imperfect repair actions are assumed to lower the system degradation through a first-order arithmetic reduction of age model.

The preventively maintained system is modeled through a semi-regenerative process with continuous state space. The kernel of the underlying Markov renewal process is calculated using martingale techniques in the spirit of [2]. Markov renewal equations are obtained for several reliability indicators. Numerical examples illustrate the behavior of the system. All details may be found in [3].

Keywords: reliability; imperfect maintenance; delayed maintenance; ARA1 model; gamma process; Markov renewal theory.

[1] L. Doyen and O. Gaudoin (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety* **84**(1): 45-56.

[2] Jean Bertoin (1996) *Lévy Processes*. Cambridge Tracts in Mathematics **21**. Cambridge University Press, Cambridge.

[3] Sophie Mercier, I.T. Castro (2013) On the modelling of imperfect repairs for a continuously monitored gamma wear process through age reduction. *Journal of Applied Probability* **50**(4), accepted publication.

A New Analytic Approximation Technique for Options in a Regime-Switching Market

Melissa Mielkie, Matt Davison

Western University Canada

Pricing and hedging options assumes first deterministic and then more sophisticated stochastic volatility stock price models. Observation of real market data suggests that volatility, while stochastic, is well modelled by moves between just two states. We propose that the transitional probabilities of volatility are given by a two-state Markov model, and that the actual jumps between volatility regimes are driven by Poisson processes. A set of coupled pricing partial differential equations (PDE) are derived using standard hedging and arbitrage arguments. Our equations are of the same form of those found in the regime-switching literature. The market price of volatility risk arises in our coupled PDE utilizing an approach based on contract independence used in previous stochastic volatility models. Using classical applied mathematics techniques, an approximate solution is derived as a function of known Black-Scholes relations since the equations cannot be solved directly. As a result, our approximate solution introduces error into our system of

pricing equations which can be quantified by considering a related root-finding problem. A comparison of our results to those found by implementing the Crank-Nicolson numerical scheme are given.

Polya - Aeppli processes in Risk Theory

Leda D.Minkova

Sofia University, Bulgaria

In this study I give a brief review of the properties and characterizations of the Polya - Aeppli process. It is discussed as a point process as well as a spacial case of a delayed renewal process. Then I continue with the bivariate version of the process with some properties and recursion formulas. The bivariate Polya - Aeppli process is a pure birth process. A bivariate insurance risk model in which the counting process is a bivariate Polya - Aeppli process is also discussed. For the bivariate Polya - Aeppli risk model the ruin probability is analyzed.

Keywords: Bivariate Polya-Aeppli process, risk model, ruin probability

Estimating distribution of age at menarche based on recall information

Sedigheh Mirzaei S., Debasis Sengupta

Applied Statistical Unit, Indian Statistical Institute, India

Average age at menarche finds application in a variety of contexts. It is a comparative indicator of population health and timing of maturation, and is also widely used as a demographic indicator of population fecundity. The most common approach of estimating age at menarche is the 'Status quo' approach, which makes use of dichotomous data and a logit/probit analysis. Dichotomous responses (whether menarche has occurred till the day of observation) are easy to obtain by asking respondent girls whether they have experience menarche, and statistical routines for logit and probit analysis of dichotomous data are available in most statistical packages. Recall data contain more information than 'status quo' data, and are expected to produce better estimates. Since recall data are generally interval censored, some scientists have used the non-parametric estimator proposed by Turnbull (1976). However, the nature of censoring involved in gathering retrospective menarchial data is informative. Alternative modelling has so far been limited to a parametric set-up. In this work we provide a non-parametric estimator, based on a likelihood that makes use of the special nature of the data at hand. Monte Carlo simulations produce encouraging results on the performance of the proposed estimator.

CALCULATING OF VALUE AT RISK IN RISK MANAGEMENT AND APPLICATIONS

Görkem MIZRAK, Sevil ŞENTÜRK, Gültekin ATALIK

Anadolu University, Faculty of Science Statistics Department, TURKEY

Risk management, the financial crisis in the international financial markets in the 1990s has gained a special importance and in terms of the companies made it necessary to measure risk accurately and more efficiently. Moreover, the impact of globalization, increasing risks due to the complex and volatile investment transactions has supported this requirement. For this reason, in a certain time period, Value at Risk which it reveal size of risk with only one number has gained importance.

In this study; Risk, Risk Management, with Value at Risk, analysing variance-covariance method in parametric method will be analyzed Monte Carlo simulation in simulation method will be accentuated. On application step of study, two applications related to both methods will be approached with real data.

Keywords: Risk Management, Value at Risk

The inhomogeneous bisexual branching process: Asymptotic growth rates

Manuel Molina, Manuel Mota, Alfonso Ramos

University of Extremadura, Spain

The bisexual Galton-Watson process in varying environments was introduced in Molina et al. (2003) as a two-type sequence $(F_n, M_n)_{n=1}^{\infty}$ defined in the form:

$$(F_{n+1}, M_{n+1}) = \sum_{i=1}^{Z_n} (f_{ni}, m_{ni}), Z_{n+1} = L(F_{n+1}, M_{n+1}), n = 0, 1, \dots \quad (1)$$

where the empty sum is considered to be $(0, 0)$, $Z_0 = N \in \mathbb{Z}^+$ and, for every $n = 0, 1, \dots$, $(f_{ni}, m_{ni})_{i=1}^{\infty}$, is a sequence of i.i.d. non negative integer valued random variables, being $p_{jk}^{(n)}$, the offspring probability distribution corresponding to the n -th generation, namely $p_{jk}^{(n)} := P(f_{n1} = j, m_{n1} = k)$, $n = 0, 1, \dots$

$(F_n, M_n)_{n=1}^{\infty}$ and its associated sequence of mating units $Z_{n=0}^{\infty}$ are Markov chains not necessarily homogeneous. We consider a superadditive inhomogeneous bisexual branching process and investigate the limiting behaviour of $F_{n=0}^{\infty}$ (respectively $M_{n=0}^{\infty}$) suitably normalized by its corresponding rates of growth.

Key words: Asymptotic behaviour, Branching process, Bisexual branching process

Comparison of Some Effect Size Measures: One-Way and Two-Way Fixed Effect ANOVA Models

Ahmet MOLLAOĞULLARI, Erkut AKKARTAL, Mehmet MENDEŞ

Canakkale Onsekiz Mart University, Faculty of Art and Science,
Department of Mathematics, Turkey

It is well known that one-way and two-way fixed effect analyses of variance (ANOVA) are used to compare differences among two or more independent group means. The differences among the means may be very large or small. However, the ANOVA does not show how the means differ from one to another. Effect size measures such as eta-squared, partial eta-squared, omega-squared and epsilon-squared help the researchers to estimate how a large difference may be found. This study was carried out to compare performances of different effect size measures namely, eta-squared, partial eta-squared, omega-squared, epsilon-squared for one-way and two-way ANOVA models. For this aim, random numbers were generated from normal (0,1), Beta (5,2), Beta (2,5) and Beta (10,10) distributions. Simulation results showed that all effect size measures were affected from sample size, distribution shape, true effect size values, variance ratio and number of groups to be compared. In general, the estimates which made by using omega-squared and epsilon-squared were closer to the real effect size values when variances were homogenous.

Keywords: Effect size, one-way ANOVA, two-way ANOVA, simulation

The market value of life insurance liabilities under a regime switching process

Rosario Monter

University of the Balearic Islands, SPAIN

This paper studies the values of the equity and liabilities of life insurance companies in the presence of regime switching in the economy. Following the contingent claim work of Grosen and Jorgensen (2002), where the equity and liability of a life insurance company are evaluated as a barrier option framework, this paper proposes a model where the dynamic evolution of the assets follows a geometric Brownian motion with parameters switching according to a continuous-time Markov chain process with discrete state values. After deriving valuation formulas, numerical implementation is illustrated using US life insurance data, providing strong evidence of switching behavior on the market affecting the contingent claim valuation.

Keywords: life insurance; regime switching; barrier options; contingent claims valuation

**The role of sociodemographic and contextual determinants:
multilevel analysis of maternal mortality in Cameroon**

MOODJOM Roland marc

Demographer and economist

Maternal health is one of the major public health problems in developing countries. The international community has set targets to reduce the level of poverty in the world through what they called the MDGs. Thus, the fifth goal focuses on maternal health. Despite decades of improvement since the health status of children in these countries, it leads us to believe that inequalities in mortality between socioeconomic groups of women (education level of parents, living in the household sector activity of parents, etc.). Different have not reduced and the mortality gap between regions.

The maternal mortality rate is a key indicator of the health of a population and development indicators. Despite all the measures taken by both the international community and the government in Cameroon, maternal mortality has increased significantly during the period 2004-2011 compared to the period 1991-1998 (EDS-MICS, 2011). What can be the sociodemographic factors and community behind this increased rate of maternal mortality.

Our fundamental purpose is based on data from the DHS in Cameroon in 2004 and 2011; show the impact of demographic status of the household and the community on maternal mortality.

In this study we attempted to examine the sociodemographic and contextual determinants of maternal mortality in order to understand why the level of maternal mortality has increased and is still very high in Cameroon. It is also to highlight the causal relationship between maternal mortality and socio-demographic status and regional disparities. To identify the complete chain of factors that influence mortality, we use a multi-level (three levels: children, household, community) to data from Demographic and Health Survey, conducted in this country. This three-tiered approach allows to deal with the dependence between the probability of death for women in the same household, village or neighborhood. In addition, it can also measure the random variations between households or between communities linked to differences in mortality among women who compose them. It therefore meets the fundamental limitations of conventional techniques of analysis based on the independence between observations without estimating the random variations due to differences between groups.

Chernoff-Based Hybrid Tau-Leap (this is a poster presentation)

Alvaro Moraes, Raul Tempone, Pedro Vilanova

King Abdullah University of Science and Technology
Saudi Arabia

Markov pure jump processes can be applied to modeling many phenomena, e.g. chemical reactions at molecular level, protein transcription and translation, dynamics of wireless communication networks, spread of epidemic diseases in small populations among others. Exact algorithms, like SSA by Gillespie or Modified Next Reaction Method by Anderson simulate a single trajectory exactly, but can be time consuming when many reactions take place during a short time interval. The approximate Gillespie's tau-leap method, on the other hand, can be used to reduce computational time, but introduces a time discretization error that may lead to non physical values. This work presents a hybrid algorithm for simulating individual trajectories which adaptively switches between the SSA and the tau-leap methods. The switching strategy is based on the comparison of the expected inter-arrival time of the SSA and an adaptive time step size derived from a Chernoff-type bound for the one-step exit probability. Since this bound is non-asymptotic we do not need to make distributional approximation for the tau-leap increments.

This hybrid method allows: (i) to control the global exit probability of a simulated trajectory,

(ii) to obtain accurate and computable estimates for the expected value of any smooth observable of the process with almost minimal computational work.

We present numerical examples that confirm the theory and show the advantages of this approach over both, the exact methods and the tau-leap ones that uses pre-leap checks based on gaussian approximations for the increments.

Time Series Analysis of Guayas Basin

Daniel Moran-Zuloaga^{1,2}, Paúl Carrión¹, Paola Romero², Garcés Daniel², Raúl Mejía³

¹Escuela Superior Politécnica del Litoral ESPOL, ²Centro de Estudios Aplicados a las Ciencias de la Tierra CIPAT, ³Instituto Nacional de Meteorología e Hidrología INAMHI, Ecuador

Guayas basin is one of the most important sources of fresh water in Ecuador and covered an area of 34500 km². The understanding of the water reservoir in this area may help to planify measurements in case of

extreme events due to Global Change. Monthly precipitation time series of the two oldest meteorological stations were analyzed in order to determine how the climate variability has affected the region surrounding the Guayas Basin. Because of the high variability in the rainfall, and several non-parametrical analyses and Auto Regressive Moving Average ARMA were performed. The results showed that after four decades and two ENSO (El Niño Southern Oscillation) events. The wettest and the driest stations were stationary after forty-three years and they were correlated each other with (Kendall, $\tau=0.62$) (Spearman, $\rho=0.80$). The results emphasize the importance of long-term data analysis. In addition, after forty-three years the rainfall followed a linear pattern. It is presumable that in the case of an ENSO event in the coming years, the rainfall average will maintain the same. In summary, this study provided us an idea of the rainfall pattern in the Guayas basin and enhances the importance of non-parametrical analysis for scattered rainfall databases.

MDG and 1000 Day Window: Health and Mortality Disparities in India

Barun Kumar Mukhopadhyay

Achieving universal access to reproductive health by 2015 is one of the two targets of *Goal 5 - Improving Maternal Health* - of the eight [Millennium Development Goals](#) (UN, 2000). The scientists, economists and health experts of the International Food Policy Research Institute (IFPRI), USA (2012) opined and agreed "improving nutrition during the critical 1,000 day window is one of the best investments to achieve lasting progress in global health and development. Considering the *Goal 5* of MDG and objective of IFPRI, an attempt is made in the paper to investigate the situation prevailing in India in relation to the health of mother, infant and children under 2 years of age. At the outset it must be mentioned that India's concern is much voiced in the international arena perhaps because of her overwhelming population of more than 1.2 billion mark (1,210,193,422 in Census, 2011). And India is likely to surpass China in population by 2030 (Rosenberg, 2012). IFPRI (2012) made some comment about India that from 1998 to 2005, India's gross domestic product (GDP) grew by 40 per cent, yet the number of children who were stunted (generally defined as being significantly below the median height for their age) declined only from 51 to 44.9 per cent, and those underweight from 42.7 to 40.4 per cent. It is very unlikely that India, having boasted of the largest democracy and second largest food producing country in the world, is still struggling to overcome the curse of ill health of mother and children, in particular, more about gender bias. It is well known that there are countries in sub-Saharan regions, in particular and other underdeveloped regions in the world, which are still clutched

with antecedent of health hazards, hunger etc. Contrary, India having glimpses of her past and present history in many fields but lag behind health, mortality, literacy, poverty and many others simply because of her heterogeneous character of people which is found less in general in other countries in the world. In this juncture, the paper tries to analyze Indian situation from different sources of data. The findings show predicaments from different major states, in general but boast of having good condition in other smaller states as well. Finally, the paper proposed a model by which people of India may be captured to know all heterogeneities rather than heterogeneities in geographical locations.

Keywords: MDG, Empowered Action Group, Index of Dissimilarity

Reach and Utilization of Maternal and Child Health Services in India

P. Murugesan

Department of Fertility Studies, International Institute for Population Sciences (IIPS), India

Maternal and Child Health has remained an integral part of the Family Welfare Programmes in India. The National Family Health Survey-3 (2005-06) provides enormous data related to Maternal and Child Health care- antenatal, delivery and post natal care for the women while for child care variables like child immunization-use of BCG, DPT 1 to 3 and Polio 1-3 etc. are available. The present paper therefore aims to provide the prevailing situation regarding reach and utilization of the MCH services in India as well as among the different states. It was observed that in India around 77 percent women received antenatal care for their most recent birth during the five years preceding the survey. Interestingly the rate of increase was higher in rural area as compared with urban areas. State-wise variations were noticed with regard reach and utilization of antenatal care services. Further, highest cases of safe delivery assisted by health personnel were found for the state of Tamil Nadu (91 percent) and lowest for Nagaland (12 percent). With regard to child care services percentage values obtained for fully vaccinated children range from 81 percent from Tamil Nadu and 21 percent for Nagaland. Besides providing state wise comparative picture with regard to MCH services the paper intends to discuss factors affecting utilization of MCH services among the regions of the country by using logistic regression analysis. In addition attempt would be made to discuss some of the plausible reasons for the variation in the utilization of MCH services.

Keywords: Reach and Utilization of Maternal and Child Health Services

Using optimized distributional parameters for continuous data modelling

K. Mwitondi¹, J. Bugrien² and K. Wang³

¹Sheffield Hallam University, Department of Computing

²Statistics Department, Benghazi University

³Radio and Space Services, Australian Bureau of Meteorology

Unsupervised and supervised modelling are the corner stone of knowledge extraction from data - the main goal of which is to identify naturally arising structures in data. One way to attaining this goal is to choose a parametric form of the density to explore the data features. The choice is generally poorly understood and any tentative choice may be too restrictive. For instance, many clustering techniques rely on the conventional mechanics of distance minimization – a feature which constitutes the very challenge they are designed to address. Further, growing volumes of data, disparate data sources and modelling techniques entail the need for model optimization via adaptability rather than comparability. We propose a novel two-stage algorithm to modelling continuous data consisting of an unsupervised stage whereby the algorithm searches through the data for optimal parameter values and a supervised stage that adapts the parameters for predictive modelling. The method is implemented on the sunspots data with inherently Gaussian distributional properties and assumed bi-modality. Optimal values separating high from lows cycles are obtained via multiple simulations. Early patterns for each recorded cycle reveal that the first 3 years provide a sufficient basis for predicting the peak. Multiple Support Vector Machine runs using repeatedly improved data parameters show that the approach yields greater accuracy and reliability than conventional approaches and provides a good basis for model selection. Multiple simulations of this type can be generated based on the algorithm above to assist in selecting the most consistent model.

Key Words: Clustering, Data Mining, Density Estimation, EM Algorithm, Sunspots, Supervised Modelling, Support Vector Machines, Unsupervised Modelling.

The cogarch models as an alternative to time series analysis: some theoretical results and applications

Ilia Negri

Department of Engineering, University of Bergamo, Italy

The COGARCH (Continuous, Generalized, Auto-Regressive Conditional Heteroschedastic) models were introduced by Kluppelberg et al. (2004) as a continuous version of the GARCH models. They are driven by general Lévy processes and the resulting volatility process satisfies a

stochastic differential equation, driven by the discrete part of the quadratic variation process of the same Lévy process. This is the main difference between COGARCH models and others stochastic volatility models which involve two sources of randomness. In this paper some recent results on inference for COGARCH(1,1) models based on two different procedures are recalled. The first is the method of moment estimation (Kluppelberg et al. 2007), the second is a prediction-based estimating functions method introduced by Sorensen (2000) and (2011) that was applied to Cogarch models by Bibbona and Negri (2013). A Monte Carlo study of the two different estimating methods is performed and the two methods are compared. An accurate analysis of financial real time series is also performed comparing the result obtained for the COGARCH models with the results for GARCH models and other stochastic volatility models.

Keywords: COGARCH Models, GARCH Models, Stochastic Volatility Models.

Replicating exchange options with transaction costs in stochastic volatility markets

Thai H. NGUYEN

University of Rouen, France

This paper studies the problem of hedging exchange options in the presence of transaction costs following Leland's approach. Adapting the form of the enlarged volatility appeared in our recent contribution cite{Nguyen}, we characterize the asymptotic behavior of the hedging error as the portfolio is frequently rebalanced. Limit theorems established enable us to obtain a super replication via a controlled parameter. A simple way to lower the option price inclusive transaction costs and some numerical illustration are also given.

Keywords: transaction costs, approximate hedging, exchange options, quantile pricing, Leland's strategy, limit theorem

Discriminant and Regression Analysis Based on Convex Piece-wise Linear Minimization with Applications to Allergological Data

Wojciech Niemiro^{1,2}, Wojciech Rejchel^{1,2}, Marta Zalewska³

¹Nicolaus Copernicus University, ²University of Warsaw, ³Medical University of Warsaw, Poland

We present algorithms based on minimization of convex piece-wise linear (CPL) functions and their applications to analysis of epidemiological data concerning prevalence of asthma in Poland.

Functions of the form $g(\theta) = \sum \rho_i(y_i - \theta'x_i)$, where $\rho_i(u) = \max(\alpha_i u, -\beta_i u)$, with $\alpha_i, \beta_i \geq 0$, are used as convenient criteria in various algorithms for statistical data analysis. In particular, Least Absolute Deviations (LAD) regression, perceptron-type classification can be obtained via minimizing such functions. Introduction of a regularizing term of the form $\sum \lambda_i |\theta_i|$ also fits in this framework and leads to variable selection method similar to well known LASSO. Moreover, this approach can be combined with the kernel trick to allow for nonlinear classification or regression. From the computational viewpoint, minimization of CPL criterion functions such as $g(\theta)$ can be done via basis-exchange techniques similar to linear programming. Specialized algorithms tailored for piece-wise linear minimization for classification and regression are efficient and reliable. Our algorithms are implemented as R functions.

We demonstrate the usefulness and flexibility of the method described above by applying several CPL-based algorithms to allergological data. One of primary goals in analysis of big data sets is selection of variables. In the course of epidemiological surveys, a large number of potentially relevant symptoms, features and pieces of information is recorded for each respondent. Only a relatively small number of them turns out to be really significant. The algorithms we use simultaneously perform feature selection and computation of classification or regression functions.

Keywords: Least Absolute Deviations, Perceptron criterion, Classification, Regression, Kernel method, Support Vector Machines, Variable selection, Allergy, Epidemiology, R.

References

1. Niemiro, W., Rejchel, W. (2009). Rank correlation estimators and their statistical properties. *Statist. Papers* 50, 887-893.
2. Bobrowski, L., Niemiro, W. (1984). A method of synthesis of linear discriminant function in the case of nonseparability. *Pattern. Reco.* 17, 205-210.

Optimal investment decision under regulatory and environmental risks

Michi Nishihara

Osaka University, Japan

This paper investigates the decision-making of a firm that has an option to invest from among multiple alternative projects. This type of option is called a max-option, and the nature of a max-option has been investigated in several papers. I extend the previous analysis to a model that allows the random occurrence and disappearance of alternative projects in which to invest. The occurrence and disappearance of opportunities in which to invest will be caused by changes in regulation, the exit and entry of rival firms, technological innovation, political risk,

catastrophes, etc. By proving the properties of the options, this paper reveals how a firm should deal with regulatory and environmental risks. Specifically, I demonstrate that the prospective future occurrence of an alternative (e.g., deregulation) has the significant effect of increasing the option value and deferring the investment decision. The results help better understand investment decisions with regulatory and environmental uncertainty.

Keywords: Finance; Decision analysis; Real option; Regulatory risk

Regression Models for Repairable Systems

Petr Novák

Charles University in Prague, Faculty of Mathematics and Physics,
Department of Probability and Mathematical Statistics, Czech Republic

When operating a device which is a subject to degradation, we want to estimate the distribution of the time to failure for maintenance optimization. Our aim is to describe the dependency of the failure time distribution on applicable regression variables. Models commonly used in survival analysis, such as the Cox model or the Accelerated failure time model, need to be adjusted to accommodate repairs and maintenance. For instance, we may use the number of repairs or maintenance actions or their cost as time-varying covariates. In this work we describe such models and demonstrate their application on real data.

Keywords: Reliability analysis, Repair models, Regression.

Modeling of stochastic hydraulic conditions of pipeline systems

N.N.Novitsky, O.V.Vanteyeva

Energy Systems Institute SB RAS, Russia

The problems of calculation of hydraulic conditions are the basic problems in the analysis of operating conditions of pipeline systems (PS) when they are designed, operated and controlled online. Solving these problems consists in finding the distribution of pressures and flow rates among the PS components under specified boundary conditions which are normally represented by flow rates of transported medium at sites of consumption. These problems are traditionally solved using deterministic models and methods which, however, do not allow one to quantitatively estimate the extent to which the operating conditions are satisfactory under uneven consumption conditions which is typical of the majority of practical cases. This situation is explained by high complexity and size of pipeline systems (water-, gas supply-, etc.) as objects to be modeled,

extreme laboriousness of the general methods for stochastic modeling (a kind of the Monte-Carlo method) as well as difficulties in obtaining primary statistical information.

Researchers of the Energy Systems Institute, Siberian Branch of the Russian Academy of Sciences, have developed an approach and a set of mathematical models, methods and techniques for modeling the pipeline system operating conditions, taking into account the stochastic character of consumption processes, their dynamics and specified control rules [1,2] which provide rational combination of modeling adequacy and computational effort.

The results presented in the paper include:

1. A technique for a priori calculation of statistical characteristics of probabilistic consumption processes, which rests on the hypothesis of normal distribution.
2. A general scheme of probabilistic calculation of PS hydraulic conditions which implies determination of statistics of their parameters at specified characteristics of boundary conditions and interaction among operating parameters which is represented by flow distribution models. It is shown that such a calculation reduces to solving a traditional problem of flow distribution [3] at the point of mathematical expectation of boundary conditions in combination with an additional procedure for calculating covariance matrices of operating parameters.
3. A technique for obtaining analytical expressions for covariance matrices of operating parameters, as well as expressions for the general case of boundary conditions specification.
4. A technique for probabilistic modeling of dynamics of hydraulic conditions on the basis of developed analytical probabilistic flow distribution models which considerably reduces computational efforts as compared to the known methods for simulation modeling.
5. A numerical example of probabilistic calculation of steady-state flow distribution in PS in comparison with the results obtained by applying the Monte-Carlo method. The example illustrates the advantages of the suggested approach.

References

1. Novitsky N.N., Vanteyeva O.V. Problems and methods of probabilistic modeling of hydraulic conditions of pipeline systems/ Nauchno-Technicheskiye Vedomosti SPbGTU, - 2008. – No.1. – P.68-75.(in Russian)
2. Novitsky N.N., Vanteyeva O.V. Modeling of stochastic flow distribution in hydraulic circuits // Izv. RAN. Energetika, - 2011, No.2. – P.145-154.(in Russian)
3. Merenkov A.P., Khasilev V.Y. Theory of hydraulic circuits. – M.: Nauka, 1985. – 280 p. (in Russian)

No data for older persons (60 years) in the Ugandan National HIV/AIDS policies, strategies and programmes

Joseph Nyende

Delta Partnership, Uganda

Background: Civil society AIDS service organizations: TASO, SAIL, UNASO, HEPS, AIC, MAMAS CLUB & NACWOLA (National HIV Prevention Advocacy Group for Older persons) are advocating for issues and data of older persons to be included in the HIV/AIDS National policies, programmes and strategies. A key challenge to this project was lack of commitment by policy and decision makers to be accountable and responsive for lack of data for older persons.

Methodology: During the review of the National Strategic Plan, October 2011, the National HIV Prevention Advocacy Group for older persons developed a strategy to engage decision/policy makers to make commitments to prioritize and include issues of HIV/AIDS for older persons and the Citizens (older persons) and CSOs-ASOs get a basis to demand for issues of HIV and AIDS for older persons in National HIV/AIDS policies, strategies and programmes.

The National HIV Prevention Advocacy Group for older persons in Uganda reviewed National HIV and AIDS policy documents to ascertain their level of inclusion of issues of older persons in those documents; the group has also reviewed the Uganda HIV/AIDS Prevention and Control Bill 2010. It also carried out a survey as a way of gathering evidence for advocacy. A survey report, policy brief and a position statement on the HIV/AIDS bill were developed and are being used for advocacy all over the country.

Results:

- Policy and decision makers are now more committed to including issues of older persons in the National HIV/AIDS Policies, strategies and programme.
- The policy reviews, survey report and position paper formed the basis for CSO-ASOs and citizens (older persons) to hold policy and decision makers accountable on issues of older persons and HIV&AIDS in the country.
- Ministry of health has committed to collecting data for people of 60 years and above during health surveys and researches carried out in the country.
- CSOs (National HIV Prevention Advocacy Group for older persons) and Policy and decision makers, Uganda AIDS Commission and government lined ministries and agencies are able to engage each other and work in partnership on issues of common interest.

Conclusions: Engagement of policy and decision makers to make commitments on inclusion of issues of older persons in National HIV policies, strategies and programmes and collecting data for persons

above 60 years proved an effective strategy for ASOs-National HIV Prevention Advocacy Group for older persons, some MPs have committed to raise issues of older persons in the HIV/AIDS bill when it is presented before parliament since they are receptive and attentive to issues from the electorate.

Parameter estimations in a special forward interest rate model

Balázs Nyul

Department of Applied Mathematics and Probability Theory, Faculty of Informatics, University of Debrecen, Hungary

We describe a special interest rate model which driven by a geometric spatial AR sheet and introduce a new type of Heath-Jarrow-Morton forward interest rate model. In this model we give the no-arbitrage criteria and we estimate parameters (for example volatility) with special samples by the maximum likelihood method. Finally we observe the asymptotic behavior of the maximum likelihood estimator in each cases.

Keywords: forward interest rate, HJM, volatility, maximum likelihood.

An empirical study on the index of satisfaction of student allocation in the Portuguese undergraduate engineering courses

Raquel Oliveira¹, A. Manuela Gonçalves², Rosa M. Vasconcelos³

¹Faculty of Social Sciences, Catholic University of Portugal, ²Department of Mathematics and Applications, CMAT - Centre of Mathematics of University of Minho, ³Department of Textile Engineering, 2C2T - Centre for Textile Science and Technology, University of Minho, Portugal

This paper aims to describe and characterize student allocation in the Portuguese public higher education system, namely in the academic engineering programs. The application of multivariate methodologies for the evaluation of the students' satisfaction index detects natural clusters of academic programs and identifies some determinants of students' choices concerning higher education. These determinants may be used as explanatory variables of a model for the access to higher education engineering programs. The data used in this paper concerns the academic year of 2010/2011 and was provided by the Portuguese Ministry of Education.

Keywords: Students' index satisfaction, higher education, education policy, clusters analysis;

Linear Regression Models with Censored Data: Bias Correction

Jesus Orbe, Vicente Núñez-Antón

Departamento Economía Aplicada III (Econometría y Estadística),
Universidad del País Vasco/Euskal Herriko Unibertsitatea, Spain

This paper focuses on regression analysis when the response variable is subject to censoring. The most frequently type of censored data is that of right censored observations. That is, the case when the observed response variable Y is such that $Y = \min(T, C)$, where C is the censoring value and T is the real value of the response variable under study. This situation is very common when analyzing duration, survival or lifetime data. Several methods have been proposed to model the effect covariates have on a right censored response variable.

The most commonly used models for this type of data can be classified into two main classes: hazard regression models and accelerated failure time regression models (see, e.g., Kalbfleisch and Prentice, 2002). Under the former approach, researchers study the effect the covariates have on the hazard function of the response variable T . Under the accelerated failure time regression models approach, the direct effect covariates have on the response variable T is studied.

In this work we concentrate on the second class of models and, more specifically, on the modelling approach proposed by Stute (1993), which belongs to the least squares regression methods class, but appropriately adapted to handle a censored response variable. This work proposes a bias reduction for the regression coefficients' estimator for linear regression models when the response variable is randomly censored and the error's probability distribution is unknown. The proposed bias correction is based on model-based bootstrap resampling techniques that also allow us to work with censored data. Our bias-corrected estimator proposal is evaluated and its behavior assessed in simulation studies, concluding that both the bias and the mean square error are reduced with the new proposal.

Keywords: Bias; Bootstrap; Censoring; Least squares.

Male-female compositional balance in the analysis of longevity for an insurance portfolio: the case of Mexican life tables

Arely Ornelas, Montserrat Guillén

Dpte. Econometria, Riskcenter-IREA, University of Barcelona, Spain

We model the mortality behavior in the Mexican general population and compare it to the mortality assumed in the life tables that are used for in the design and analysis of insurance products that involve some longevity risk. We describe the data from 1990 to 2009 and model mortality dynamics with standard demographic models. We also adjust a Brass-type model to compare the gap between the general population mortality rate estimates and the mortality rate estimates for the insured population that are being used by the National Insurance and Finance Commission in Mexico. Some life tables currently used by insurance companies are not different for male and female. We assume possible scenarios of gender proportion in the insurance mortality tables. We compare our results for Mexico with the ones obtained for the population in Switzerland and we find very similar results. We emphasize the limitations of unisex mortality tables for insurance and, in general, the incurred bias in risk estimation if the proportion of male and female is not balanced in the portfolio, the same way that it is in the life table. We will to present a simulation study that shows the implications of neglecting the effect of male-female composition in the analysis of longevity for an insurance portfolio.

Keywords: Mortality rates, Lee Carter, Longevity dynamics, Insured population

Comparison of multi-stage dose-response mixture models, with applications

Eva Maria Ortega¹, José Alonso²

¹RESID. UNIVERSITARIA OBLTS., SPAIN

²Hospital Virgen Caridad, Cartagena, Spain

This article is concerned with the stochastic comparisons of the population who generate a response within the multi-stage hit model by introducing random dose. We consider the multistage hit mixture model for the population generating a response, whose mixing distribution is given by the dose and the size of the population at risk (or a vector of other exposure parameters), that leads to random sums of mixed bernoulli random variables. Unlike the previous literature, we deal with parameter uncertainty with arbitrary mixing distribution. A uniform

distribution can be adjusted to model the dose. For the population generating a response, we state conditions for the increasing convex order, the increasing concave order, the likelihood ratio order, the stochastic order, the hazard rate order, the reversed hazard rate order and the mean residual life order. We compare the multi-stage hit mixture model with positively correlated parameters, and the multi-stage hit model with independent parameters, via the variability of the response, using the concept of positive quadrant dependence. Using an univariate ageing concept, the bounds of the response can be calculated from the models with fixed parameters having an exact distribution. These results highlight the effect of positive correlations and variation of parameters to assess the level of magnitude and the degree of variability, exhibited by the response to random dose.

Keywords: dose-response, multi-stage hit models, mixed bernoulli, stochastic orderings, variability, mixtures, bounds, medicine

Comparisons of multistate models with discrete-time pure-birth process for recurrent events and uncertain parameters

EM. Ortega¹, E. Ortega-Quiles², J. Alonso³

¹RESID. UNIVERSITARIA OBLTS., SPAIN

²Colegio Hnas. Carmelitas, Elche, Spain

³Hospital Virgen Caridad, Cartagena, Spain

This paper presents a multistate model with recurrent events that are modeled by a discrete-time pure birth process. The probability distribution of the multistate model is based on random sums with support on the natural numbers and with geometric summands. Exact distributions are given in Belzunce, Ortega and Ruiz (2009) for independent inter-arrival times. These models are analyzed by stochastic orderings in the case of independent geometric, arbitrarily distributed, and positively correlated inter-arrival times. Some multivariate extensions are described. Motivations are provided in reliability, economics, biology and demography. We also provide results for some qualitative properties of the kind 'new better than used' of the multistate model assuming iid summands, that are illustrated.

Keywords: multistate models, discrete-time pure birth process, recurrent events, unobserved heterogeneity, stochastic orderings, ageing notion

Mathematical Modeling (ROR) applied to the forecast of earthquakes in the global level

Osés Rodríguez, Ricardo, I Pedraza Martínez, Félix Alfredo, Saura González, Guillermo

Centro Meteorológico Provincial Villa Clara, Cuba

Objective: Modeling the data series of global earthquakes that have occurred on Earth during the period from 1990 to 2010 and the total number of estimated deaths due to these phenomena. **Methods:** Objective Regression Methodology named ROR.. **Results:** We obtained correlation coefficients between actual and forecast for total earthquake (0.987) and 0.890 for all deaths, with errors and 38235.9 2853.33 respectively, both variables depend on the value of six years ago, the trend the total number of earthquakes is increasing annually in 2119.98 also performed a forecast through 2016 observed that the number of earthquakes increase if the trend holds up, the highest value is presented in 2015 and in 2016 provided a slight decrease to the number of deaths is an increase in 2014 and in 2016 it should be noted that only the bodies of the Civil Defense are responsible for issuing alerts on the occurrence of these phenomena and this work is part of studios Alternative methods for predicting the amount of deterministic earthquakes. **Discussion:** Our work is consistent with similar work in our province but modeled data series of major earthquakes that have occurred in the national territory of Cuba in the period from 1551 to 1992, both using the same methodology differing from this work in that in this model year, month, day, time, latitude and longitude, and the magnitude of next earthquake, we see these promising results in terms of modeling the total amount of earthquakes and the forecast number of deaths caused by them.

Conclusions: The tendency of all earthquakes is increasing in 2119.98 annually, the number of earthquakes increase if the trend holds up, the highest value is presented in 2015 and in 2016 is expected to decline, the number of deaths has an increase in 2014 and in 2016,

Keywords: Global Modeling | forecast | Earthquakes | deaths.

Optimal Insurance with Investment and Loans

Darya Ostrovskaya

Department of Probability Theory,
Lomonosov Moscow State University, Russia

The aim of this paper is to construct an optimal strategy for an insurance company. The company is supposed to be able to invest its capital or to borrow with some interest rate r . A company with initial capital x operates during n periods, in the beginning of each period it gets premiums equal to c and makes a decision about the value invested or borrowed y . In the end of the period company gets claims from the clients η_i , which are independent identically distributed random variables with the distribution function $F(x)$ and density $f(x)$. If the insurer does not have enough capital

to pay the claims, then he borrows lacking money immediately at a higher rate q . The aim of the company is to find the optimal value of investment (or borrowing) y^* in order to minimize expected costs. Optimal strategy is investigated in several discrete-time insurance models with the aid of dynamic programming.

Let $G(x,y)$ be expected costs for one period, let $fn(x)$ be expected costs for n periods provided that in each period choice of y was optimal. Hence we may derive Bellman equation $fn(x)=\min_y[G(x,y)+Efn-1(P(x,y,\eta))]$, where $P(x,y,\eta)$ is the capital of the company in the end of one period with initial capital x , value of investment (borrowing) y and claims η for this period. In case of equal rates for investment and borrowing optimal solution of this equation is found explicitly: $y^*(x)=F-1((q-r)/q)-x-c$. Moreover, the model with different rates for investment and borrowing is also investigated.

For obtained solutions global sensitivity indices of the parameters of the model are computed using Sobol' decomposition.

Estimating child mortality from information on previous birth: data from a Portuguese birth cohort

Neir A. Paes¹; Cristina Teixeira²; Henrique Barros³

¹Department of Statistics of the Federal University of Paraíba, Cidade Universitária, Brazil, ^{2,3}Institute of Public Health, University of Porto, Porto, Portugal and Department of Clinical Epidemiology, Predictive Medicine and Public Health, University of Porto Medical School, Porto, Portugal

A whole range of techniques have been developed for estimating infant and child mortality from the information routinely recorded in maternity registers on age and reproductive history of delivering mothers. Among them, the best-known and most widely applied indirect technique was developed by William Brass. Using the baseline information from a birth cohort (Generation XXI) assembled in the period 2005-2006 in the metropolitan area of Porto, North of Portugal, we addressed in this paper the following objectives: (1) to estimate the child mortality rates based on information about previous births and their survivorship among women recruited for this birth cohort; (2) to compare these indirect estimates with direct values retrieved from Portuguese vital statistics; (3) to analyse the potential of this birth cohort to generate plausible estimates of life-table indicators. We retrieved data on mother's age, previous live births prior to the current one, and number of surviving and deceased children from a group of multiparous women ($n=3521$). The data was divided into seven 5-year groups by maternal age and survival and death probabilities were computed for each group. Through the Brass method,

we obtained estimates of probability of dying before attaining certain exact childhood ages, $q(x)$, by using the multipliers $k(i)$ as proposed by Palloni-Heligman. Then, a logit life-table system was used to derive life-table indicators. Accordingly, probabilities of dying between birth and 2, 3, 5 and 10 years were respectively: 4.0; 5.7; 7.3 and 9.6 per 1,000 children ever born, which were allocated in time-period. These indirect estimates compared with the direct ones obtained from Portuguese vital statistics revealed that they were very similar. The life expectancy at birth was 77.6 years for both sexes, and the implied infant mortality was 4.0 per 1000 live births for Grand Porto during the period 2005-2006. The use of indirect method to analyze the potential of the Generation XXI cohort data in provide relevant information on reproductive issues, not available in the vital statistics, seems to be an important and effective tool, and promissory for analyzing the follow-up studies of this cohort held in 2009 and 2012.

Keywords: Child mortality, Brass' Method, Preceding Birth Technique.

Probabilistic Approach to Clustering in Stochastic and Agent-Based Computational Models

Florentin Paladi, Ghennadii Gubceac

Department of Theoretical Physics, Faculty of Physics and Engineering,
State University of Moldova, Moldova

A theoretical model which describes the interactions in a heterogeneous population of agents was developed and analyzed within the framework of the discussions about the gap between agent-based computational models (ABM) and stochastic analytical models. The herding and clustering of agents in complex systems is a typical problem dealt with new approaches, for example, to macroeconomic modeling, that describe macroscopic variables in terms of the behavior of a large collection of microeconomic entities. By definition, an ABM model looks at a system not at the aggregate level which could possibly be described with a few equations of motions, but at the level of its constituent units, and this microscopic level modeling includes describing the individual behavior of potentially many agents in the system. Simulating them can be extremely computation intensive and time consuming. It might not be possible to carry out exponentially increasing number of such simulations within the framework of a traditional ABM model, regardless of the computing power implied, simply because it is impossible to explore the whole set of interactions for the applied strategies. The number of possible partitions for N agents into m subsets or clusters can be easily generalized, and all values for the average payoffs, and variances are computed. Finally, we show that even a relatively simple probabilistic model can describe precisely the expected outcomes from

corresponding agent-based simulations. In general, we want to get new insights into microscopic explanations of stochastic models which can be compared with the agent-based computational models, and so forth to bridge the gap between ABM models and stochastic processes. The model can be used to study, in particular, individual interactions in heterogeneous populations, as well as the structural relaxations in the complex systems based on multi-agent interaction concept.

Keywords: Complex System, Stochastic Process, ABM, Cluster.

Optimal quantization and transmission for ECG signals using genetic algorithms

A. Pastor-Sánchez, E. Alexandre-Cortizo, S. Salcedo-Sanz, D. Gallo-Marazuela, L. Cuadra-Rodríguez

Department of Signal Theory and Communications, Escuela Politécnica Superior, Universidad de Alcalá, Spain

In this paper an algorithm for optimal Electrocardiogram (ECG) signal quantization driven by a genetic algorithm (GA) is proposed. The purpose of this quantization method is to minimize the Percentage Root Mean-square-difference (PRD) of the quantized signal. The aim of this genetic algorithm is to choose the optimal number of bits to quantize the ECG, given the maximum number of available bits and the maximum distortion allowed.

A novel GA-based ECG coding system will be proposed, using the above-mentioned quantization algorithm. The system will automatically select the optimum frame size, by means of an additional evolutionary algorithm, and, for each frame it will transmit the quantized values jointly with the step size used for that frame. The proposed method will be tested and can be used to efficiently store or transmit ECG signals.

Multistage Non-Homogeneous Markov Chain Modeling of the Non-Homogeneous Genetic Algorithm and Convergence Results

André G. C. Pereira, Viviane S. M. Campos

Mathematics Department, Universidade Federal do Rio Grande do Norte, Brazil

In this work, the results presented in the paper titled Multistage Markov Chain Modeling of the Genetic Algorithm and Convergence Results are extended. The size of the neighborhood in the crossover stage of the Algorithm 1, is permitted to vary throughout the evolution of the algorithm and instead of a fixed mutation probability, in mutation stage of the Algorithm 2, a sequence of mutation probabilities is introduced. Finally,

numerical simulations are developed and comparisons between the original algorithms and the new versions of them are presented.

Keywords: Simulated annealing, Global optimization, Genetic Algorithms

Modeling of Mortality in Elderly Women due to Cardiovascular Disease in the Northeast of Brazil

Jozemar Pereira dos Santos¹, Neir Antunes Paes²

¹Department of Statistics, Federal University of Paraíba, João Pessoa, PB, Brazil, PhD Candidate in Health Sciences, UFRN, Natal, RN,

²Health and Decision Modeling Postgraduate, Federal University of Paraíba, Brazil

An ecological study was conducted in 187 micro regions in the northeast of Brazil to compare two types of statistical models used to identify the effect of living conditions on the mortality rate due to cardiovascular disease (CVD) in elderly women. For this study, the population data were taken from the Brazilian 2000 Census and the mortality data from SIM/MS (Mortality Information System, from the Brazilian Health Ministry). Two statistical models were compared; in model-1 the outcome variable Mortality Rate due to CVD was directly observed and in model-2 it was indirectly estimated from four potential causal factors. The Structural Equation Modeling (SEM) used showed better fit to model-1, with significance in the measurement model for the following indicators: Years of study, Percentage of elderly women living in homes with bathroom and running water, Probability of living to age 60; and fit indicators of the structural model: Discrepancy Function $\chi^2_{(g,l=2)}=1714$ significant p-value = 0,424 (indicating absolute fit); Normed Chi-square, $\chi^2/(g.l)=0,857$; Goodness-of-fit Index, GFI=0,995; Adjusted Goodness-of-fit Index, AGFI=0,976; Normed Fit Index, NFI=0,989; Tukey-Lewis Index, TLI=1,000; Root Mean Square Error of Approximation, RMSEA=0,000.

Keywords: Aged. Cardiovascular Disease. Causality. Factor Analysis. Mortality. Social Conditions

EXAMINING THE FACTORS OF SUCCESS OF THE FOOTBALL TEAMS IN 2011-2012 SUPER LEAGUE SEASON IN TURKEY VIA PLS PATH MODELING

Esra Polat, Semra Turkan, Suleyman Gunay

Hacettepe University, Faculty of Science, Department of Statistics, Turkey

Partial Least Squares (PLS) is a family of multivariate data analysis methods that allows for analyzing blocks of variables over a set of

observations. The variety of PLS methods depends on the number of blocks of variables and the relationships among blocks. The common denominator of all PLS methods is the fact that they are based on an iterative algorithm in which the parameters are calculated by a series of least squares regressions. Due to the introduction of PLS, one of the main Structural Equation Modeling (SEM) techniques, by Wold in 1966, it has been received with considerable interest among researchers. In most cases, the PLS approach to SEM is usually named as PLS Path Modeling (PLS-PM). PLS-PM is a multivariate data analysis methodology, which provides a framework for analyzing multiple relationships between a set of blocks of variables. It is supposed that the relationships among the blocks are established taking into account previous knowledge (theory) of the phenomenon under analysis. In the context of PLS-PM, latent variables (LVs) are unobserved variables or variables that cannot be measured directly, for instance, beliefs, intention and motivation. Hence, they are measured indirectly via manifest variables (MVs) (indicators, items) which could be perfectly observed-measured. Every PLS-PM is formed by two sub models: the structural (inner) model and the measurement (outer) model. The structural model is the part of the model that has to do with the relationships between the LVs. In turn, the measurement model is the part of the model that has to do with the relationships of a LV with its block of MVs. The purpose of this study to examine the success of 18 Turkish national football teams in 2011-2012 Super League Season by using PLS-PM. For this purpose, following the study of Sanchez (2013), a model is proposed in which the overall success of the football teams depends on the quality of the attack as well as on the quality of the defense made by them. There are three LVs which are defined as attack, defense and success. The number of won matches at home and the number of won matches away variables are taken as indicators of success, the number of goals scores at home and the number of goals scores away variables are taken as indicators of attack and the number of goals conceded at home and the number of goals conceded away are taken as indicators of defense.

Keywords: Football Teams, Attack, Defense, Success, PLS, PLS-PM.

What do student satisfaction questionnaires measure?

Fabiola Portillo¹ and Cecilio Mar Molinero²

¹Department of Economics and Business, University of La Rioja, Spain

²University of Kent, UK, and Universitat Autònoma de Barcelona, Spain

University students are asked to complete satisfaction questionnaires in which express their opinion on many aspects of their studies. It is hard to think what purpose such questionnaires may serve: it is very difficult to dismiss university teacher on the basis of students' perception of their performance; and staff are promoted on the basis of the success of their

research rather than on the quality of their teaching. From a quality enhancement point of view, one is led to conclude that student satisfaction surveys, by using valuable time staff time, produce more harm than good. And yet, there are many important questions that could be answered with the help of student questionnaires. One obvious example is whether teaching should be done by research active staff or by staff who do no research and specialise in teaching methods. Another aspect that is the impact on students' perception of "sexing up" a subject. This would militate against intellectually challenging subjects, such as those that contain advanced quantitative analysis. Finally, one can ponder on the question of class size. Up to what point is the student's perception of the teaching quality influenced by personal contact?

In this study we collected summary information on questionnaire results for all the undergraduate modules for which such information was available in a management school in a major UK university for the academic year 2008/2009. We also collected information on average marks achieved in the same modules, as well as a set of contextual variables. The management school offered 56 undergraduate modules but, following a computer crash, information on student questionnaires was lost for some of the modules, and full data is only available for 45 modules. 37 lecturers were involved in the teaching. An estimate of research activity was also available for each lecturer, having been developed as a by-product of the 2008 research assessment exercise of British universities.

The data was analysed using various multivariate statistical tools, including multiple regression and principal components analysis. It was found that students' perception of teaching quality was not related to the research rating of the lecturer, and that the only factors that influenced it were class size and mathematical content. It was also found that an index of teaching quality developed from the questionnaires was not related to results achieved.

Modelling financial data using distributions tailored on given moments of the empirical distribution

V. Poti^a, M. G. Zoia^b, M. Faliva^b,

^aDublin University, Ireland, ^bUniversità Cattolica di Milano, Italia

Risk measures, including Conditional Value-at-Risk (Expected Shortfall), turn out to be quite sensitive to the degree to which distributions are thick tailed and asymmetric. Lack of encoding information about asymmetry, leptokurtosis and non-linear dependence is a well-known drawback of the Gaussian law. This, on the one hand, has led to a search for

alternative distributions (Student t, Pearson type VII, normal inverse Gaussian, several stable distributions, see, e.g., Mills *ibid.*, Rachev et al. 2010), and, on the other hand, has accelerated the development of approaches based on copula theory and related techniques (Nelsen 1999, Szegő 2004). In this paper, we will tackle the issue of accounting for asymmetry, (possibly severe) excess kurtosis and dependence by following the alternative approach of adjusting bell-shaped distributions using orthogonal polynomials as shape adapters.

Our focus will be on polynomial transformations of parent symmetric probability density functions to match the empirical moments of target distributions characterized by possibly substantial heavy-tails and asymmetry. We will demonstrate a simple but powerful novel result, i.e. that the coefficients of the polynomial that achieves the transformation are simple algebraic functions of the difference between the moments of the target distribution and those of the parent distribution. We will then apply this result to the modelling of heavy-tailed and skewed distributions of financial asset returns.

In a recent paper (Zoja, 2010), the focus was on the Gaussian law as the parent distribution with Hermite polynomials as shape-adapters, showing how the normal distribution can be tailored to suit excess kurtosis (up to 4), skewness as well as between-squares correlation. In this paper, the focus will be instead on the logistic distribution. The leading argument for the choice of the logistic distribution as the parent distribution is that it is by itself leptokurtic (its kurtosis being equal to 4.2, that is 40% higher than that of the Gaussian) and, what is more, the kurtosis can be further pushed up to 15.44 by an orthogonal polynomial shape adapter. The possibility of encoding severe excess kurtosis, together with asymmetry and between-squares dependence, via orthogonal-polynomial techniques, makes the logistic law a valuable candidate for the modelling of the distribution of financial asset returns..

In a generalization to the multivariate case, we will then present an alternative approach to the use of copulas in modelling dependencies among multiple stochastic processes, e.g. the returns on the assets included in an investment portfolio, and will apply our results in a GARCH setting, to allow for distributions of conditional innovations with realistically fat tails and skewness without the need to resort to poorly globally identified and hence hard to estimate stochastic volatility and jump models.

Keywords: Orthogonal Polynomials vs. Moments, Kurtosis, Asymmetry, Between-squares Dependence.

Mathematics Subject Classification 33C45; 62E10

Forecasting wind farm financial return

F. Prattico¹, G. D'Amico², F. Petroni³

¹Dipartimento di Ingegneria Industriale e dell'Informazione e di Economia, Università degli studi dell'Aquila, ²Dipartimento di Farmacia, Università "G. D'Annunzio" di Chieti-Pescara, ³Dipartimento di Scienze Economiche ed Aziendali, Università degli studi di Cagliari, Italy

The variations of the wind speed in a certain site are strictly correlated to the economic aspect of a wind farm, for example maintenance operations especially in the offshore farms, pitch angle control on new wind turbines or also evaluation of new specific sites. To better evaluate the amount of energy produced, a good model for wind speed is needed. In a previous paper we introduce an indexed semi-Markov process that is able to reproduce the most important statistical features of wind speed, namely the probability density function and the autocorrelation function. In this work we use that model in order to forecast the wind speed and its related energy in a specific site. We do forecasting one step ahead for a long period and for different time scale. The wind speed and the energy produced, real and forecasted, are compared through common statistical indicators.

Keywords: Semi-Markov, wind speed, wind energy, forecasting.

Modeling survival data using Lindley-Geometric distribution and some extensions

Vasile Preda¹, Silvia Dedu²

¹ Faculty of Mathematics and Computer Science, University of Bucharest, Romania

² Department of Applied Mathematics, Bucharest University of Economic Studies, Romania

We propose some extensions of Lindley-Geometric distribution for modeling survival data. Several properties of these extended distributions are investigated. Maximum likelihood estimations of the parameters of the new distributions are derived. Computational results are provided.

Keywords: Lindley-Geometric distribution; Survival data; Mixed distribution; Parameter estimation.

Acknowledgement. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-RU-TE-2012-3-0007.

On series $\sum c_k f(n_k x)$ with random frequencies

Marko Raseta, Istvan Berkes

Graz University of Technology, Austria

There is a wide and nearly complete theory of trigonometric series with random coefficients, but there exist only scattered results on trigonometric series with random frequencies. As it turned out in recent years, such series are very useful for constructions in uniform distribution theory and the theory of lacunary series.

$\sum c_k f(n_k x)$ in the case when f is a periodic measurable function and n_k is an increasing random walk, i.e. the gaps $n_{k+1} - n_k$, $k=1,2,\dots$ are i.i.d. positive random variables. In particular, we prove a Strassen type functional law of the iterated logarithm for sums $\sum c_k f(n_k x)$ with periodic Lipschitz functions f . We obtain an analogous result for empirical distribution function of $n_k x$, $1 \leq k \leq N$, describing its limiting behavior in terms of reproducing kernel Hilbert spaces. This yields, in particular, the precise asymptotics of the discrepancy of $n_k x$, a difficult open problem in the case of nonrandom n_k .

Keywords: trigonometric series, law of the iterated logarithm

Confidence Intervals for Dynamic Concentration/Inequality Indices of Economic Systems Modeled by Birth-Death Processes

Philippe Regnault

Laboratoire de Mathématiques de Reims, UFR Sciences Exactes et Naturelles, Moulin de la Housse, BP1039, 51687 REIMS, France

In this paper, we are interested in estimating a dynamic index of the concentration of wealth shared by a population of economic agents – Theil's entropy. At any time, each agent produces an amount of wealth, independent of other agents. According to this wealth production, a wealth class (among predefined classes) is associated to each agent. We suppose that wealth class allocation is governed by a continuous-time birth-death process so that the repartition of total wealth's production among classes at any time is an explicit function of birth and death rates. Then, by considering Theil's entropy of wealth distribution among all classes at any time yields dynamic concentration index of wealth. If moreover, the birth-death process is ergodic, then Theil's entropy associated to the asymptotic repartition of wealth measures the

concentration of wealth when agents' population reaches its equilibrium state.

When the class allocation process is observed at equidistant discrete times, it is necessary to estimate Theil's entropy index. In the same fashion as in Regnault [10], we prove the strong consistency and asymptotic normality of plug-in estimators of dynamic and asymptotic Theil's entropy, built from empirical estimators of the transition matrix of the discretized class allocation process. Confidence intervals for dynamical Theil's entropy are then derived.

Keywords: Population Dynamic, birth-death Process, Theil's Entropy, Reward, Estimation.

Confidence Intervals for Dynamical Concentration/Inequality Indices of Economic Systems modeled by Continuous-time Markov Processes

Philippe Regnault¹, Guglielmo D'Amico²

¹Laboratoire de Mathématiques de Reims, UFR Sciences Exactes et Naturelles, FRANCE

² Università di Chieti-Pescara Chieti Italy

In this talk we are interested in estimating some dynamical indices of the concentration of wealth shared by a population of economical agents. At each time, each agent produces an amount of wealth, independently of the other agents. A wealth class (among predefined classes) is allocated to each agent according to his wealth production. We suppose that the class allocation is governed by a continuous-time birth-death process so that the repartition of total wealth production among classes at any time is an explicit function of birth and death rates. Then, by considering the Gini index and Theil's entropy of the wealth distribution among all classes at any time yields dynamical concentration indices of wealth. They constitute stochastic versions of the classical Gini index and Theil's entropy; see D'Amico and Di Biase (2010, 2012) for details. If, moreover, the birth-death process is ergodic, then both the Gini index and Theil's entropy associated to the asymptotic repartition of wealth (the asymptotic Gini and Theil's entropy indices) measure the concentration of wealth when the population of agents reaches its equilibrium state.

When the class allocation process is observed at equidistant discrete times, it is necessary to estimate Gini and Theil's entropy indices.

In the same way as in Regnault (2011), we prove the strong consistency and asymptotic normality of plug-in estimators of both the dynamical and asymptotic Gini and Theil's entropy indices, built from empirical

estimators of birth and death rates. Confidence intervals for the trajectories of Gini and Theil's entropy processes are then derived.

Keywords: Birth-death processes, Confidence interval, Dynamic economic systems, Estimation, Gini index, Theil's entropy.

Kernel and ESIS estimates of J-divergence

Martin Řezáč

Department of Mathematics and statistics, Faculty of Science, Masaryk University, Czech Republic

J-divergence is one of the widely used indexes (beside Gini index and K-S statistic) for assessment of predictive models. It is defined by integral of transformed conditional densities of output of given model, where bivariate target variable is modeled. Generally, the J-divergence is the symetrized Kullback-Leibler divergence.

Commonly it is computed by discretisation of data into bins using deciles with requirement on the nonzero number of cases for all bins. Two alternative methods to this approach can be used. First, it is the kernel smoothing theory, which allows estimating unknown densities and consequently, using some numerical method for integration, to estimate value of the J-divergence. ESIS estimator is the second alternative.

The main objective of the contribution lies in providing a simulation study showing properties of considered estimators. Specifically, it is focused on the bias and mean squared error (MSE) of the estimators according to size of data sample, selected distribution types and selected parameters of given data sample.

Keywords: J-divergence, Kernel smoothing, ESIS.

A partial review of notions of positive dependence, relations and applications

Yosef Rinott

Hebrew University of Jerusalem and LUISS, Italy

This paper is a review of the more useful notions of dependence: Association, MTP₂ and related notions and inequalities, Quadrant dependence, Regression dependence, and related stochastic orderings. I will then discuss some applications in probability and statistics, such as limit theorems, bounds on significance levels (FDR), and more.

Longitudinal Analysis of renal function with informative observation times

Lisandra Rocha¹, Inês Sousa^{1;2}, Raquel Menezes^{1;2}

¹Center of Mathematics of University of Minho, Portugal,

²Department of Mathematics and Applications of University of Minho, Portugal

In this work, we intend to analyze the progression of renal function in patients of Hope Hospital in Salford, England. The aim of this work is to model the evolution of creatinine in the blood taking into account that time measurements depend on the responses obtained. For that, we propose a model that considers dependence between the longitudinal outcome process and the follow-up time process. This model use subject-specific random effects to describe the association between the two processes of interest and a joint-likelihood-based estimation was developed for estimation. We will also analyze the same data with a standard longitudinal model which ignore the dependence assumption, aiming to compare the results of the two analyzes. The principal difference between these two models and different approaches is that, the standard longitudinal model considers the follow up time process as deterministic while the model that we propose considers the follow up time process stochastic.

Keywords: Follow up Time Process, Longitudinal Analysis, Monitoring Kidney Function.

Statistical estimation of non-stationary MAP2s

Joanna Rodríguez, Rosa E. Lillo, Pepa Ramírez-Cobo

Universidad Carlos III de Madrid, Spain

Since Markovian arrival processes (MAP) were introduced in the eighties they have received much attention in different contexts from teletraffic or finance to queueing. Their capability of modeling dependent and non-exponentially distributed observations makes the MAPs a versatile class of stochastic models.

A number of articles have considered statistical estimation for the MAPs, but always under the assumption that the process is in its stationary version, which may be restrictive in practice. In this work we address inference for the two-state nonstationary MAP given a set of multiple samples. Our procedure is based on the combination of a moments matching technique with a maximum likelihood approach.

Keywords: Markovian Arrival Processes, Maximum likelihood estimation, Phase Type processes.

Uncertainty in Predicted Mortality Rate

Klemens Rosin, Cornelia Schwierz, Judith Riegelnig

Zurich Statistical Office, Switzerland

Zurich – Switzerland's largest city – has considerably changed its age structure over the last thirty years. During this period mortality rates have decreased substantially. Observed past trends are incorporated in prediction models to project human mortality until 2025. This study evaluates different model types based on their model prediction uncertainty following generalized likelihood uncertainty estimators.

First, parsimonious model approaches (e.g. linlog) are compared with standard models (e.g. Gompertz). Second, different model stratification levels (gender, nationality, city district, profession, marital status) and their interactions are tested with respect to influence on prediction uncertainty. Third, effects of considering meteorological events (temperature drop, extreme low and high temperature, solar radiation) in the models are evaluated. On the whole, this study presents a framework how to assess parsimonious human mortality models based on prediction uncertainty to achieve robust human mortality predictions.

Keywords: mortality forecasts, model selection, theories of uncertainty

Wrong assessment of mortality from diabetes in Russia

TP Sabgayda, DO Roschin, VG Semyonova

Federal Research Institute for Health Organization and Informatics of
Ministry of Health and Social Development of Russian Federation,
department of the analysis of health statistics, Russia

Russia is among ten countries with the largest number of diabetic patients and with one of the lowest recorded levels of mortality (0.45% in deaths structure in 2010). That indicates existence of disorders in registration of causes of death. Purpose: to identify the reason of extreme underreporting of mortality from diabetes mellitus. The main reason for the low rate of mortality is that the death causes from the complications of diabetes recorded as other classes of causes. But this factor is common for all countries. While in Russia the share of unreported deaths from diabetes is too large. The next factors were analyzed: the completeness of recorded diabetes in patients with in-hospital mortality, the share of records about diabetes as a concomitant disease in death certificates, the frequency of use of the code E14 (unspecified diabetes) for registration of death as well as codes that can not be the primary cause of death, and to be replaced by diabetes. The result shows that the errors made in coding the causes of death as well as the taking into account the only underlying cause of death in statistics are the determinants of underestimation of mortality from diabetes in Russia.

Keywords: encoding rules, underlying cause of death, undiagnosed diabetes, in death certificates, share of diabetes in the deaths structure

Description of gene expression from microarray data by fractional stable distributions

Saenko Viacheslav, Saenko Yuriy

Ulyanovsk State University, Russia

Statistical investigations of gene expression had been carried out in last some years show the non-Gaussian character of gene expression profile. These PDFs possess of kind typical properties. First these PDFs are universal and they describe gene expression both bacteria and human. The tails of these distributions decrease by power law [1]. Therefore it were used for describing of such kind of distribution both Zipf-Pareto distributions [2,3] and log-normal distribution [4].

In this work the fractional-stable distributions are used for description of gene expression profiles. This kind of distributions belong to the class of limit distributions and possesses all properties mentioned above. The PDFs of fractional-stable distribution is

$$q(x; \alpha, \beta, 0) = \int_0^{\infty} g(xt^{\beta/\alpha}; \alpha, 0) t^{\beta/\alpha} g(t; \beta, 1) dt,$$

where $g(x; \alpha, 0)$ is symmetric stable distribution and $g(x; \alpha, 1)$ is one-sided stable distribution with characteristic function

$$\hat{g}(\lambda; \alpha, \theta) = \exp\left(-|\lambda|^\alpha \exp\left(-i\left(\frac{\pi}{2}\right)\theta \alpha \operatorname{sign} \lambda\right)\right)$$

The parameters of the distributions were statistically estimated with respect to experimental sample.

The algorithm of statistical estimation of the parameters will described in the work. This algorithm is applied to statistical estimation of the parameters of PDFs for gene expression of various genes. A comparison of PDFs of gene expression of various genes and physical reason of such behavior of fluorescent marks will present.

The work is done by support Ministry of Education and Science of the Russian Federation (grant No. 14.B37.21.0558)

Keywords: fractional-stable distribution, gene expression micro-matrix array

[1] H. R. Ueda, S. Hayashi, et. al, Universality and flexibility in gene expression from bacteria to human, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 11, 3765–9, 2004.

[2] C. Lu and R. D. King, An investigation into the population abundance distribution of mRNAs, proteins, and metabolites in biological systems, *Bioinformatics (Oxford, England)*, 25, 16, 2020–7, 2009.

- [3] D. C. Hoyle, M. Rattray, R. Jupp, and A. Brass, Making sense of microarray data distributions, *Bioinformatics (Oxford, England)*, 18, 4, 576–84, 2002.
- [4] T. Konishi, Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment, *BMC bioinformatics*, 5, 5, 2004.
- [5] V. E. Bening, V. Y. Korolev, et. al, Fractionally stable distributions, in *Stochastic Models of Structural Plasma Turbulence*, V. Y. Korolev and N. N. Skvortsova, Eds. Utrecht: Brill Academic Publishers, 175–244, 2006.

Statistical Estimation Based on Generalized Order Statistics from Kumaraswamy Distribution

Samir K. Safi

Department of Statistics, The Islamic University of Gaza, Palestine

The Kumaraswamy distribution is similar to the Beta distribution but has the key advantage of a closed-form cumulative distribution function. In this paper we present the estimation of Kumaraswamy distribution parameters based on Generalized Order Statistics (GOS) using Maximum Likelihood Estimators (MLE). We proved that the parameters estimation for Kumaraswamy distribution can not be obtained in explicit forms, and therefore it has been implemented using the simulated data for illustrative purposes. We compare the performances of parameters estimation through an extensive numerical simulation for different sample sizes. These simulations examine the sensitivity of estimation to different sample sizes. In particular, how do estimations perform for small, moderate and large sample sizes? The main findings are: First, the worst performance estimation for small sample size selection for different values of the parameters estimation. Secondly, as the sample size increases the MSE of the estimation decreases. Finally, the estimation accuracy reaches its superiority for large sample sizes.

Keywords: Kumaraswamy Distribution, Generalized Order Statistics, Simulation, Maximum Likelihood Estimators.

The Coral Reefs Optimization Algorithm: An Efficient Meta-heuristic for Solving Hard Optimization Problems

S. Salcedo-Sanz, J. Del Ser, I. Landa-Torres, S.Gil-López and J. A. Portilla-Figueras

Department of Signal Theory and Communications, Escuela Politécnica Superior, Universidad de Alcalá, Spain

This paper presents a novel bio-inspired algorithm to tackle complex optimization problems: the Coral Reefs Optimization (CRO) algorithm.

The CRO algorithm artificially simulates a coral reef, where different corals (namely, solutions to the optimization problem considered) grow and reproduce in coral colonies, fighting by choking out other corals for space in the reef. This fight for space, along with the specific characteristics of the corals' reproduction, produces a robust meta-heuristic algorithm, shown to be powerful for solving hard optimization problems. In this research the CRO algorithm is detailed and tested in several continuous and discrete optimization problems, obtaining advantages over other existing meta-heuristic techniques. The obtained results confirm the excellent performance of the proposed algorithm.

Estimating the Density and Hazard Rate Functions Using the Reciprocal Inverse Gaussian Kernel

Raid B. Salha

Department of Mathematics, The Islamic University, Palestine

In this paper, we use the Reciprocal Inverse Gaussian (RIG) kernel to estimate nonparametrically the probability density function (pdf) and the hazard rate function for independent and identically distributed (iid) data. The estimator uses adaptive weights depending on the points at which we estimate the functions. We derive the strong consistency, the asymptotic normality and the asymptotic mean squared error (AMSE) of the proposed estimator. Also, the selection of the optimal bandwidth is investigated. The performance of the proposed estimator is compared to that of the Gaussian kernel.

Keywords: Reciprocal Inverse Gaussian kernel, hazard rate function, kernel estimation, asymptotic mean squared error, boundary bias.

2000 MSC: 62G07, 62G08

Incorporating the Stochastic Process Setup in Parameter Estimation

Lino Sant, Mark Anthony Caruana

Department of Statistics and Operations Research, Faculty of Science,
University of Malta, Malta

Estimation problems within the context of stochastic processes are usually studied with the help of statistical asymptotic theory and proposed estimators are tested with the use of simulated data. For processes with stationary increments it is customary to use differenced

time series, treating them as selections from the increments' distribution. Though distributionally correct, this approach throws away most information related to the stochastic process setup. In this paper we consider the above problems with reference to parameter estimation of a gamma process. Using the derived bridge processes we propose estimators whose properties we investigate in contrast to the gamma-increments MLE. The proposed estimators have a smaller bias, comparable variance and offer a look at the time-evolution of the parameter estimation. Empirical results are presented.

Keywords: Lévy processes, gamma process, bridge process, Dirichlet distribution.

Migration Flows Estimation Using The Modified Gravity Model

Diana Santalova

Institute of Mathematical Statistics, University of Tartu, Estonia

In the present research the international and/or domestic migration flows are estimated using the modified gravity model. The modified gravity model is a nonlinear regression model for estimating symmetric passenger correspondences between pairs of spatial points [1]. Unknown parameters of the model and correspondences are estimated using aggregated data, i.e. total number of passenger departures from each point in a considered time interval is used. The method of estimation under the assumption that aggregated data and the error term are distributed normally is carried out in [1-2]. However non-symmetry of migration flows may violate the assumption about normality of errors distribution. We intend to change the model with normally distributed errors to a model with skew-normal error distribution, and to propose a method of its estimation.

Keywords: gravity model, migration flows, skew-normal distribution

References:

- [1] Andronov A., Santalova D. Statistical estimates for modified gravity model by aggregated data. *Communications in Statistics - Simulation and Computation*, Taylor & Francis, 2012, 41:6, pp. 730-745.
- [2] Santalova D. Experimental Analysis of Statistical Properties of Parameter Estimation of Modified Gravity Model. *Automatic Control and Computer Science*, 2013, 47:2, in print.

Defining informative prior distributions for stochastic model estimation: a Chaos Polynomial-based approach

Francisco J. Santonja, Ana Corberán-Vallet

Department of Statistics and Operations Research, Universitat de València, Spain

A key aspect in the analysis of a stochastic model from a Bayesian viewpoint is the choice of the prior probability distributions of the model parameters. Our focus here is to propose a Chaos Polynomial-based approach to define informative priors, which will be fundamental when the available information is limited.

Chaos Polynomial conception considers that any random variable, $X(\omega)$, in an Hilbert space can be defined as a polynomial series expansion in a standard normal variable ; that is $X(\omega) = \sum_{i=0}^{\infty} a_i H_i(\xi(\omega))$, where $\{H_i(x), i \in \mathbb{N}\}$ are the Hermite polynomials.

An estimation of the polynomial chaos expansion coefficients a_i associated with each of the model parameters using maximum likelihood allows us to know the first and the second moments of their distributions. These moments will be considered to define the informative prior distributions. In order to show this approach, we will calibrate an epidemiological-type model.

Keywords: Chaos Polynomial, Bayesian Stochastic models, Prior distributions.

Changes in Causes of Death in the City of Zurich

Cornelia Schwierz, Klemens Rosin, Judith Riegelnig

Zurich Statistical Office, Switzerland

Data on causes of death (COD) provide information on mortality patterns and form a salient element of public health information. During recent decades the causes of death and mortality rates in Zurich – Switzerland's largest city- have undergone considerable change. During the same period, Zurich's population has become younger and better educated.

We consider COD in the time period from 1995-2011. The aims of this study are a) to describe the spatial patterns at the city district level and temporal evolution of the COD in Zurich, b) to identify to which degree the spatio-temporal changes in COD can be explained by socio-economic (income, educational status, profession), environmental (temperature, air quality) and demographic (age, gender, nationality, marital status) covariates, and c) to estimate to which degree the

resulting statistical models for COD and mortality can be incorporated in the current population projection model of Zurich to improve the projected mortality rates for the period until 2025. The study has important ramifications for city planning in the sectors health, care and geriatrics

Keywords: causes of death, spatio-temporal patterns, statistical models

Stochastic correlation in financial markets

Luis Seco

University of Toronto, Department of Mathematics, Canada

This will be an expository account of stochastic correlation models in financial markets, which will also include some recent results. The models are discussed in the context of asset management and aim at an enhanced portfolio theory where the Gaussian assumptions are relaxed.

Keywords: Portfolio optimization, stochastic correlation, mathematical finance

Approximation of a Random Process with Variable Smoothness

Oleg Seleznev¹, Enkelejd Hashorva², Mikhail Lifshits³

¹Department of Mathematics and Mathematical Statistics, Umeå University, Sweden, ²Actuarial Department, Faculty HEC, University of Lausanne, Switzerland, ³Department of Mathematics and Mechanics, St.Petersburg State University, Russia

We consider the problem of approximation of a locally stationary process with a variable smoothness index. An example of such process is a multifractional Brownian motion, which is an extension of the fractional Brownian motion with path regularity varying in time. These models are applied, for example, in image processing, finance, biomedicine, and geosciences. Assuming that the smoothness index attains its unique minimum in the interval (an isolated singularity point), we propose a method for construction of observation points sets (sampling designs) in piecewise constant approximation. For such method, we find the exact asymptotic rate for the integrated mean square error. Further, we show that the suggested rate is optimal, e.g., convergence is faster than for conventional regular designs. The obtained results can be used in various problems in signal processing, e.g., in optimization of compressing digitized signals, in numerical analysis of random functions, e.g., in simulation studies with controlled accuracy for functionals on realizations of random processes.

Keywords: variable smoothness, multifractional Brownian motion, piecewise constant approximation.

Identification of a Simple Homeostasis Stochastic Model Based on Active Principle of Adaptation

Innokentiy V. Semushin, Julia V. Tsyganova, Anatoli G. Skovikov

Ulyanovsk State University, Russia

The Active Principle of Adaptation for linear time-invariant state-space stochastic MIMO filter systems is applied to human body temperature daily variation adaptive stochastic modeling.

Keywords: Adaptation, active principle, homeostasis, parameter estimation, stochastic modeling, thermoregulation.

The distorting death causes structure of Russian population

V. Semyonova, T. Sabgayda

Federal Research Institute for Health Organization and Informatics of
Ministry of Health and Social Development of Russian Federation,
Russia

The analysis of codes of death causes was conducted on the base of official mortality database of 2010. The next factors distorting death causes structure of working population have been revealed: the factors which lead to unsatisfactory quality of diagnostics of death causes in total (insufficient frequency of post-mortem examinations; there is no opposition to over the frequent use of unspecified causes among medical community; lack of control of errors in analysis of logic of development lethal disease; lack of examination of patients before their deaths; low qualification of medical specialists in the sphere of death causes coding); the main factors which lead to worsening of quality of mortality statistics from external causes (unsettled legal framework; tendency in forensic medical practice to use the codes R96-R99 instead of Y10-Y34; non-compliance to unified standards of death causes coding among all Russian regions); the obstacles hindering objective registration of alcoholic component of mortality including cardiovascular diseases (establishment of alcohol-related diagnosis together with narcologist only; long-standing practice of establishing of alcohol-related diagnosis only when there is no other alternative; filling of the line in death medical certificate "other important conditions which contributed death" is not obligatory).

Keywords: quality of mortality statistics; reliability of reduction of mortality from external causes; reliability of growth of cardiovascular mortality; unspecified diagnosis; underlying death cause.

**Registration of vital events and verbal autopsy in rural area.
Experience of Millennium Villages Project in Potou site
(SENEGAL): Challenges and Weakness**

Massamba SENE

Health Coordinator of Millennium Villages Project, Potou/SENEGAL

In sub-Saharan Africa, the majority of people die without having the opportunity of being seen by a doctor before death, nor an autopsy that could have been done after their death, particularly in rural areas. The causes of death are difficult to study in this part of the world owing to this lack of data. The impact of interventions to reduce mortality in a community cannot be really measured if we stick only to data collected in health structures. The verbal autopsy allows us to determine the causes of deaths attributed to a specific disease and the measures to be taken to prevent such deaths; its results allow us to define priorities and target our interventions. A comprehensive registration of vital events is a crucial step in the investigation of community health problems. Not only it provides precise information on the mortality rate in a given population, but it is also the foundation of all investigations towards formulating the priorities and planning the corresponding interventions. In the context of the Millennium Villages Project (MVP), the impacts of the interventions towards reaching the MDG 4 and 5 cannot be measured without an in-depth understanding of the under-five and maternal mortality.

1/ Weakness: The basic method to record births and deaths is done through household visits (HV) by community health workers (CHW). These HV are often the only occasion in a month to account for vital events at the household level. However, this method does not accurately reflect the situation for two main reasons:

- ✓ **The inadequate household coverage:** though each household should be visited at least once a month, it is possible that the households are not visited throughout the month because of an inadequate coverage.
- ✓ **The frequency of the visits:** even with an adequate coverage, the households are only routinely visited once a month.

Those two reasons are responsible for the missed opportunities in reporting vital events.

2/ Challenges: In order to find solutions to the operational problems related to the CHW's work, the Potou site has identified alternatives sources of information. Those sources are identified at different levels of the continuum of care (community, health center and referral structures). The coordinator of the CHWs and the Verbal Autopsy Specialist crosscheck the information collected from these different sources in order to delete duplicates.

- ✓ At the community level: the main challenge is to engage some community leaders such as the Imams, Village Chiefs, members of women empowerment group, the network for young girls and the "Badienou gohk"². These leaders can be used as CHWs in the inventory of vital events at the community level. They can therefore share this information with the CHWs when the latter are doing their regular visits. The community radio station and the toll free number can also be used to share this information.
- ✓ At the health center level: the involvement of the health professional is important due to their crucial role in the continuum of care.
- ✓ At the referral structure level: information are obtained from the feedback provided by the assistant-nurse in evacuations who provide a monthly report on the issues and conditions of evacuations

Involvement of the community and religious leaders

The MVP health staff has organized a caravan in the different health posts to meet the representatives and religious leaders of the communities covered by every health post. The general objective of these meetings is to improve the census of the vital events through a greater involvement of the leaders.

More specifically the meetings with the Imams (religious leaders) and village chiefs have enabled:

- I. To raise the leaders' awareness on the impact of the census of vital events on health interventions and the importance of their involvement in the process
- II. To highlight the terms of their involvement by a close partnership with the community staff for the vital events census.

A meeting has been held in every health post with the Nurse Head of Health Post, the MVP staff and the community staff involved in the census (CHWs for CC+ and health supervisors), the community birth registration officer, the imams and the religious chiefs of the health post operating zone. The meetings were highly appreciated by the community and religious leaders and they consider that they have a key position in the program thanks to their social responsibility: they are always the first to be inform for whatever activity which is to be held in their respective communities. Important contributions were registered during the meeting, and this has enabled to set up a partnership framework:

- ✓ To provide every village chief with a register to record all vital events in the community.
- ✓ To give every village chief the telephone numbers of the supervisor, CHWs, ambulance and toll free for an early alert in case of birth delivery for the postnatal consultation to be carried out during the first week.
- ✓ Necessity of the relay to meet the village chiefs whenever they come for the regularly visits.
- ✓ To organized a restitution meeting in every village by the leaders to inform and raise the population awareness on the importance of the vital events census.
- ✓ Birth registration on the birth registers set in place must lead to a birth certificate at the registry office
- ✓ The involvement of the parent is important as well, because birth declaration is a must.
- ✓ To ensure the sustainability of the vital events census system by empowering the village chiefs in the management of the registers.

Optimal control of systems with several replenishment sources

Ksenia Shakhgildyan¹

¹ Ksenia Shakhgildyan, Lomonosov Moscow State University,
Department of Probability Theory

A model of inventory management with several sources of replenishment is examined. To study any real processes or systems we need to select an appropriate mathematical model. At first, we assume that all the parameters of the model and distributions of the underlying processes

are known. Applying the method of dynamic programming we obtain an optimal control for any planning horizon.

Let c_1 be the price of a unit ordered at the first supplier and we assume that delivery is made immediately. In its turn, the second supplier makes delivery immediately with probability p and at the beginning of the next period with probability $q=1-p$. Denote by c_2 the corresponding unit price. At the beginning of each period (a day, a week, a month, etc.) the decision to order a certain amount of goods from the first and second supplier is made, namely, $x_1 \geq 0$ and $x_2 \geq 0$ respectively. We also consider the storage cost h and shortage penalty for unit price r .

Suppose that x is the initial stock and claims are received periodically, namely, the amount ξ_i is demanded during the i -th period, $i \geq 0$. We assume that ξ_i , $i \geq 0$, form a sequence of mutually independent random variables with a common distribution function $F(\cdot)$ having density $\varphi(s) > 0$ for $s \in [a, b]$, where $a \geq 0$.

Denote by $f_n(x)$ minimum expected discounted costs over n periods. One period expected costs are equal to $L(v) = E[r(\xi_1 - v)^+ + h(v - \xi_1)^+]$, where $v = x + x_1$.

$$c_1$$

Put $G_n(u, v) = (-c_2)v + c_2 + pL(u) + qL(v) + \alpha E f_{n-1}(u - \xi_1)$, where $u = v + x_2$ and α is the discount factor.

According to the Bellman's optimality principle for $n \geq 1$ we obtain the

$$c_1$$

following recurrence relations: $f_n(x) = -x + \min_{x_1 \leq v \leq u} G_n(u, v)$, where $f_0(x) = 0$.

Note that the parameters of the model u and v corresponding to the minimum costs mean that it is optimal to order the amount $v - x$ of goods from the first supplier and the amount $u - v$ from the second.

Hence the optimal strategy of the company can be determined, namely, the values of orders at any step of the multi-step model for various values of the parameters are obtained. However, the information about the parameters as well as about distribution functions is often

incomplete. This is the reason why we examine the sensitivity of solutions to small changes in parameters and find statistical estimates of distribution functions.

Stability Selection and Randomization in L_1 Quantile Regression

Sidi Zakari Ibrahim¹, Mkhadri Abdallah¹, N'Guessan Assi²

¹Department of Mathematics, Faculty of Sciences Semlalia, Cadi Ayyad University, Morocco, ²Department of Mathematics University of Sciences and Technologies of Lille, France

Statistical models in linear regression generally focus on estimation and interpretation of conditional mean effects. However, in some situations considering mean effects could be not appropriate when for example we have great variations in response variable percentiles or when we have outliers. We here propose the Stability Selection method for variables selection in high dimension penalized linear Quantile Regression. This approach combines subsampling and variable selection algorithms adapted to the case of high dimension. Particularly, we apply Stability Selection with Lasso and Randomized Lasso Quantile Regression. Finally, the proposed method is compared with its competitors on simulated and real data sets.

Keywords: Quantile Regression, High Dimension, Resampling, Stability Selection.

Computational Algorithms for Moments of Accumulated Markov and Semi-Markov Rewards

Dmitrii Silvestrov¹, Raimondo Manca², Evelina Silvestrova³

¹Stockholm University, Sweden, ²University of Rome "La Sapienza", Italy, ³Mälardalen University, Sweden

Power moments for accumulated rewards defined on Markov and semi-Markov chains are studied. A model with mixed time-space termination of reward accumulation is considered for inhomogeneous in time rewards and Markov chains. Characterization of power moments as minimal solutions of recurrence system of linear equations, sufficient conditions for finiteness of these moments and upper bounds for them, expressed in terms of so-called test functions, are given. Backward

recurrence algorithms for funding of power moments of accumulated rewards and various time-space truncation approximations reducing dimension of the corresponding recurrence relations are described.

Alternative assessments of the probability of death with a case study for persons with Celiac Disease in selected Easter European countries

Ondrej Simpach

Department of Demography, Faculty of Informatics and Statistics,
University of Economics in Prague, Czech Republic

The probability of death and with it the hope of survival depended in the past to a considerable extent on the level of advancement of the health service, the medical findings acquired and knowledge of the appropriate treatment processes. In the case of persons with Celiac Disease, which is disease involving gluten intolerance, the hope of survival in the majority of countries was slim until the eighties of last century. These people died at a very young age thanks to ignorance of the diagnosis of their disease. However, as soon as it was possible to determine the diagnosis of Celiac Disease correctly there was a considerable breakthrough and progress rapidly changed the hope of survival for these people. This breakthrough occurred earlier in some countries and later in others. In this way treatment procedures were found for hitherto unknown diseases, or at least there was information on reducing the consequences of these diseases. The submitted study will provide a look at the alternative assessment of the probability of death of persons with Celiac Disease and the probability of death in general. The modeling of the probability of death of persons with Celiac Disease and persons in the general population is possible with the use of the LOGIT and PROBIT models of discrete selection, which are gaining considerable popularity at present in such applications as marketing, banking and insurance. On the basis of supplementary information about the population it is then possible to construct various probability scenarios with the utilization of alternative variables.

Keywords: probability of death, celiac, LOGIT, PROBIT.

Analysis of a system with three dissimilar units under pre-emptive repeat repair policy attended by two repairmen

S. B. Singh, Maneesha Gahtori

Department of Mathematics, Statistics and Computer Science, G. B.
Pant University of Agriculture & Technology, India

In the present paper we have considered a system with three dissimilar units namely A, B and C (with different failure rates). The system is in

operational state if either all units are working perfectly or the unit C and at least one of the units A or B are working efficiently. The considered system is repairable and two repairmen having different skills and efficiency are available for repairing. The subunit A has got priority to repair over the subunit B and pre-emptive repeat repair policy is followed. When both the repairmen are involved in repair jointly with different failure rates, copula technique is applied to obtain the reliability measures of the complex system. The model has been solved using supplementary variable technique, Laplace transformation and Gumbel-Hougaard family of copula. Different reliability measures of the system like availabilities, reliability, asymptotic behaviour, MTTF and cost effectiveness have been obtained. Numerical examples are then studied in detail to demonstrate the theoretical results developed in the paper.

Keywords: Reliability, MTTF, Asymptotic behaviour, Repairman, Pre-emptive repeat.

Trading with Hidden Markov Models

I. Róbert Sipos, János Levendovszky

Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary

In this paper a Hidden Markov Model (HMM) based prediction algorithm will be introduced for algorithmic trading on different financial assets and instruments. We use various training methods (e.g. Baum-Welch expectation maximization, simulated annealing and hybrid methods) for optimizing the parameters of HMM in order to capture the underlying characteristics of the financial time series. The new hybrid algorithm combines simulated annealing (SA) with the Baum Welch algorithm (used as a local search after each step of SA) and can provide relatively fast and good quality solutions. The real time nature of learning can be guaranteed by running SA only for a limited number steps determined by a predefined time interval. The algorithms are tested on US SWAP rates and FOREX series. The results demonstrate that a good average return can be obtained by HMM based prediction. It is noteworthy that the continuous model gives better result, however it requires more complex training.

The work reported in the paper has been partly developed in the framework of the project: "Talent care and cultivation in the scientific workshops of BME" project. This project is supported by the grant TÁMOP - 4.2.2.B-10/1--2010-0009.

Keywords: Hidden Markov Models, financial time series, algorithmic trading

Healthy Life Expectancy in UK: Estimates and Comments on the UK Health Performance

Christos H Skiadas¹ and Charilaos Skiadas²

¹ManLab, Technical University of Crete

²Department of Mathematics and Computer Science, Hanover College,
Indiana, USA

This note focus on the estimation of the Healthy Life Expectancy at Birth (HLEB or HALE) in UK based on a methodology presented and applied in [1]. The task is to estimate HALE from the population and death data or from the life table data sets without using the survey methods. The main advantage of the new methodology is the estimation of HALE from the same data sets needed for the estimation of the Life Expectancy at Birth (LEB) thus providing estimates for HALE for the same time periods. The data from 1922 to 2010 for UK are used and the results are summarized in the next Figures 1,2. The LEB and the HLEB without severe, severe and moderate and without all types of disability causes are presented. There are no health surveys covering all the periods studied. However, data from a recent paper published in Lancet [2] are used for a comparative study (more information in [3,4,5])..

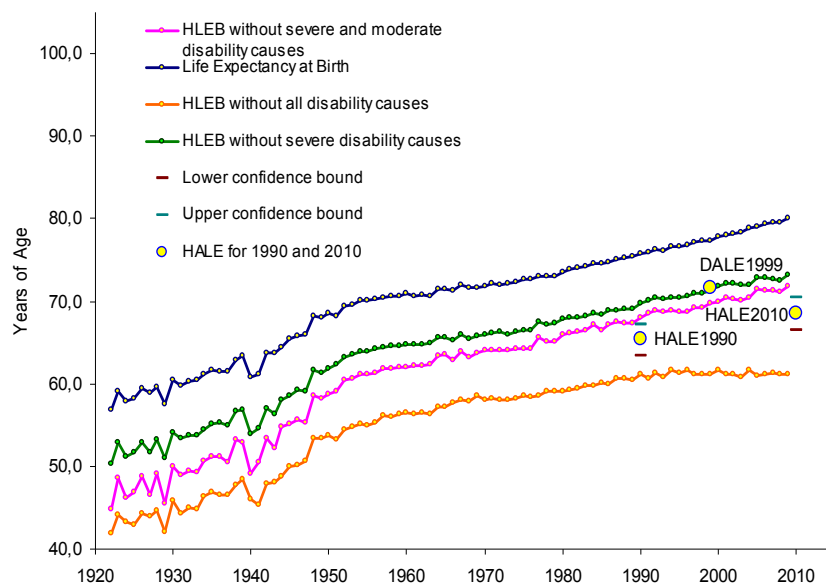


Figure 1. HALE and LEB in UK

This study provides the healthy life expectancy (HALE) at birth for the years 1990 and 2010 for UK. Both estimates are located in the interval between the HLEB without severe and moderate disability causes and the HLEB without all types of disability causes. Instead the estimate for

another form of healthy life expectancy (DALE) in 1999 for UK published in Lancet [3] is perfectly located in the line of HLEB without severe disability causes

Interesting are the comparisons of the figures suggested by the official statistics of UK [5] for males and females in the country with our results. The statistics cover the period from 1981 to 2010. Next Figure 2 illustrates our estimates for LEB and HLEB without severe, severe and moderate and without all types of disability causes for males in UK (1922-2009).

The Healthy Life Expectancy (HLE) was estimated from the “General Household Survey (GHS) question used to calculate good and fairly good general health rates: Over the last 12 months would you say your health has on the whole been good, fairly good, or not good?”

The Disability Free Life Expectancy (DFLE) was estimated from the “General Household Survey (GHS) question used to calculate free from limiting long-standing illness rates: Do you have any long-standing illness, disability or infirmity? By long-standing I mean anything that has troubled you over a period of time or that is likely to affect you over a period of time. If 'Yes': a) What is the matter with you? b) Does this illness or disability (do any of these illnesses or disabilities) limit your activities in any way? “.

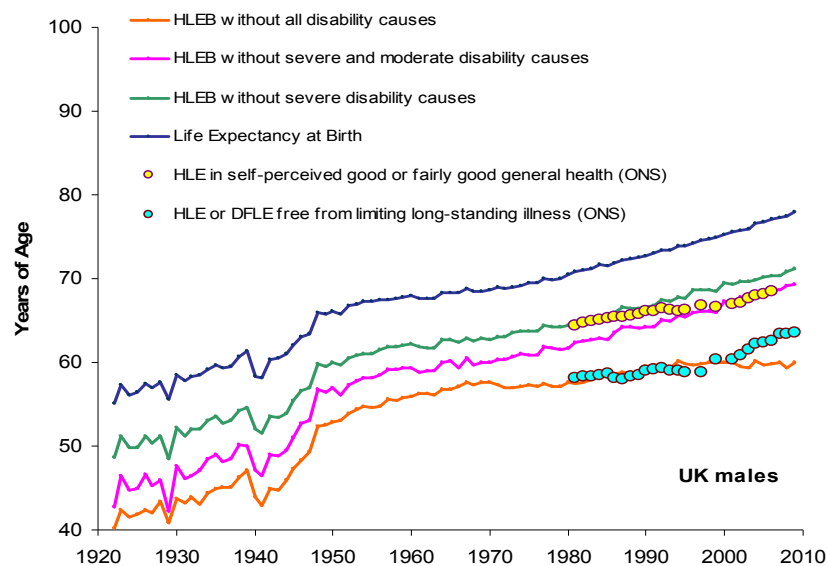


Figure 2. HALE and LEB for UK (males).

The survey methods used and the methods of analysis have changed from the first period of estimates (1981-2001) and the second (2000-2009). Even more the estimates for 1981-2001 are for Great Britain thus slightly differing from the estimates for UK. Instead the estimates for (2000-2009) are for UK. We keep the GB estimates for 1981-2001 as the

provided estimates for the second period 2001-2009 do not provide significant differences between GB and UK for both males and females as is presented in the estimates provided by the ONS Excel Chart [6](UK_healthexpectancies2000200_tcm77-229272).

In both cases we observe that the survey estimates are within the intervals estimated with our method. The Healthy Life Expectancy free from limiting long-standing illness (LLI) (1981 – 2001) fit almost perfectly on our HLEB without all disability causes, whereas the Disability Free Life Expectancy (DFLE) estimated from 2001-2009 gradually diverges from our estimates. The Healthy Life Expectancy in self-perceived good or fairly good general health 1981 – 2001 is in accordance to our HLEB without severe disability causes for males whereas for the period 2001-2009 fits perfectly to our HLEB without severe and moderate disability causes.

References

- [1] C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*, (January 2013) http://www.amazon.com/Health-State-Function-Population/dp/6188046505/ref=sr_1_1?s=books&ie=UTF8&qid=1364343495&sr=1-1
- [2] Joshua A Salomon*, Haidong Wang, Michael K Freeman, Theo Vos, Abraham D Flaxman, Alan D Lopez, Christopher J L Murray, Healthy life expectancy for 187 countries, 1990–2010: a systematic analysis for the Global Burden Disease Study 2010, *Lancet* 2012; 380: 2144–62.
- [3] Colin D Mathers, Ritu Sadana, Joshua A Salomon, Christopher JL Murray, Alan D Lopez, Healthy life expectancy in 191 countries, 1999, (World Health Report 2000) *Lancet* 2001; 357: 1685–91.
- [4] Robine JM, Mathers C, Brouard N. Trends and differentials in disability-free life expectancy: concepts, methods and findings. In: Caselli G, Lopez AD, eds. *Health and mortality among elderly populations*. Clarendon Press: Oxford, 1996; 182–201.
- [5] Office of the National Statistics (ONS) (UK_healthexpectancies2000200_tcm77-229272).
- [6] ONS Excel Chart (UK_healthexpectancies2000200_tcm77-229272).

An Interesting Inversion Problem

Christos H Skiadas¹ and Charilaos Skiadas²

¹ManLab, Technical University of Crete, Chania, Crete, Greece

²Department of Mathematics and Computer Science, Hanover College, Indiana, USA

In many first exit or hitting time cases when a stochastic process hits for the first time a barrier we face the following inverse problem: Given the probability density function $g(t)$ for the first exit time of the form (k is a constant):

$$g(t) = \frac{k}{\sqrt{t^3}} e^{-\frac{(H_t)^2}{2t}}$$

Find the unknown function H_t given the distribution function $g(t)$ from:

$$H_t = \pm \left| \left(-2t \ln \frac{g(t)\sqrt{t^3}}{k} \right)^{1/2} \right|$$

Clearly the estimation of parameter k and then of H_t is very difficult or impossible in several cases. In Demography Data (Life Table Data, Population or Mortality data sets) we have succeeded in finding a methodology to estimate k and consequently to find the unknown function H_t . The method was tested in the case of a simple process expressed by the Inverse Gaussian where H_t is a straight line and then in more complex forms. The main finding was that for the Demography data the appropriate parameter $k=k_{max}$, where k_{max} is the maximum value

of $k(t)$ provided from the form $k(t) = g(t)\sqrt{t^3}$.

Finding the form of the unknown function H_t is of particular importance in the study of many cases resulting from the first exit time of a stochastic process from a barrier that is death process systems. In this case H_t is the function expressing the way by which this system declines and can be termed as "The Health State Function of the System"[1].

Reference

- [1] C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*, (January 2013) http://www.amazon.com/Health-State-Function-Population/dp/6188046505/ref=sr_1_1?s=books&ie=UTF8&qid=1364343495&sr=1-1

Optimal Sample Size Allocation for Multi-level Stress Testing with Extreme Value Regression Under Type-I Censoring

So, H.Y., Ng, H.K.T., Gao, W.

Department of Statistics, The Chinese University of Hong Kong, Hong Kong

In this article, we discuss the optimal allocation problem in a multi-level accelerated life testing experiment under Type-I censoring when an extreme-value regression model is used for statistical analysis. We derive the expected Fisher information and the asymptotic variance-covariance matrix of the maximum likelihood estimators. Three optimality criteria are considered and the corresponding optimal allocations are determined. Under Type-I censoring, because the optimal allocations are

depending on the model parameters, the sensitivity of the optimal allocations due to mis-specification of the model parameters is studied. A numerical example is used to illustrate the methodologies developed in the manuscript.

Keywords: Accelerated Life-testing; A-optimality; D-optimality; Fisher information; Maximum likelihood estimators; Optimal design

Numerical approximation of solutions of stochastic differential equations driven by multifractional Brownian motion

Anna Soós

Faculty of Mathematics and Computer Science, Babes-Bolyai University, Romania

We study the numerical approximation of Ito stochastic differential equations driven by multifractional Brownian motion.

We consider the following stochastic differential equation driven by multifractional Brownian motion

$$X(t) = X_0 + \int_0^t F(X(s), s) ds + \int_0^t G(X(s), s) dB(s), \quad t \in [0, T]. \quad (1)$$

We assume that with probability 1 we have $F \in C(\mathbb{R}^n \times [0, T], \mathbb{R}^n)$, $G \in$

$C^1(\mathbb{R}^n \times [0, T], \mathbb{R}^n)$ and for each $t \in [0, T]$ the functions $F(\cdot, t)$, $\frac{\partial G(\cdot, t)}{\partial x}$,

$\frac{\partial G(\cdot, t)}{\partial t}$

, are locally Lipschitz. B is a multifractional Brownian motion.

The equation (1) will be approximated for each $N \geq N$ through

$$X_N(t) = X_0 + \int_0^t F(X_N(s), s) ds + \int_0^t G(X_N(s), s) dB_N(s). \quad (2)$$

We will show that the equation (2) has a local solution, which converges in probability to the solution of (1) in the interval, where the solutions exist. We use power series expansions for multifractional Brownian motion.

Keywords: multifractional Brownian motion, stochastic differential equations, series expansion.

Interpolation methods for internet traffic

Soos Anna, Somogyi Ildiko

Babes-Bolyai University, Faculty of Mathematics and Computer Science, Romania

The classical methods of data interpolation can be generalized with fractal interpolation. Our aim is to make some comparison of the fractal and numerical analysis interpolation methods. The experimental data

regarding the internet traffic were processed using fractal interpolation and also spline and some Shepard type interpolation.

Keywords: interpolation methods, fractal functions.

Cluster Analysis of Italian Revenue Agency Provincial Directorates

Andrea Spingola

Agenzia delle Entrate, Italy

This study aims to analyze the Provincial Directorates (PD), their territory and the demand of the market, building a new map of Italy not in terms of management, but of "affinities". These "affinities" are defined on the basis of a series of forty variables, selected among those which, in relation to the institutional activity of the Agency, represent both territory and market demand of the PD. The need to perform this analysis arises from the basic assumption that markets with similar characteristics involves similar risks and opportunities. In line with that assumption, the country has been divided into groups of PD with as homogeneous as possible characteristics within each group and significantly different characteristics between one group and another. The territorial analysis, highlighting the key "needs" of the catchment area, provides a synthetic and operational overview, useful for assessing the potential impacts on the Agency strategies (eg focus on the supply system and make it consistent with the demand).

This study was carried out on the basis of a multidimensional analysis to identify homogeneous groups (clusters) of PD. The main difficulty of this work, and its main advantage, is that it has reduced the complexity of a forty variables analysis.

Keywords: Multivariate data analysis, Factor analysis, Principal components analysis, Cluster analysis.

Intergenerational transmission of education in Greece: evidence from the European Social Survey 2002-2010

G. Stamatopoulou, M. Symeonaki, K. Michalopoulou

Department of Social Policy, Panteion University of Social and Political Sciences, Greece

The present paper explores the patterns of intergenerational educational mobility in Greece and its changes for different birth cohorts born between 1930 and 1976. More specifically, we investigate the transmission of educational attainments from both father and mother through generations and over time. The main purpose is to trace the transitions of individuals between educational categories and to determine the relationship between an individual's education class and the class of his or her parents. Based on data drawn from ESS (2002{2010), our analysis provides Markov transition probability matrices

and the absolute and relative mobility rates, by comparing the different rounds of the survey.

Keywords: Transmission of education, intergenerational mobility, transition probability matrices, mobility indices, cohort analysis, ESS.

Statistical estimation in discrete-time multitype branching processes – an approach based on the trimmed likelihood

Vessela Stoimenova

Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics, Department of Probability, Operations Research and Statistics, Bulgaria

Branching processes form an important class of stochastic processes with numerous applications in different scientific and practical areas. Many real phenomena of reproduction and transformation of different objects in physics, chemistry, biology involve a multitype setting. Statistical estimation of its characteristics is an important issue in their study. In our work we consider an approach for estimation, combining the asymptotic distributions of the relative frequencies of the processes, the estimation of the mean vectors and the covariance matrix of the process and focusing on the trimmed likelihood estimation as a method robust to the presence of outliers and lack of a priori satisfied theoretical assumptions. Finally we consider numerical results for simulated data and a real example of estimation of a multitype branching process arising in cell biology.

Keywords: Multitype branching processes, relative frequencies, covariance matrix, trimmed likelihood.

The Production Volume of the Piggery and the Piglets' Ability to Grow Under Favourable and Unfavourable Circumstances

Marijan Sviben

Freelance Consultant, Croatia

It has been reported that after the end of 1991 because of the war the number of weaned piglets decreased to 0.8139 in 1992, 0.7229 in 1993, 0.7298 in 1994, 0.7528 in 1995, 0.4887 in 1996 and in 1997 to 0.4593 of the number registered at the Pig Farm Nova Topola, Bosnia and Herzegovina, in 1991 (102,137). It has been established that during the war and transitional years the numbers of weaned piglets decreased because of damaging of the sows' ability to conceive and to bear the progeny, because of damaging of the sows' prolificacy too but not because of damaging the piglets' liveability. At the same place the numbers of kilogrammes of weaned piglets a year decreased more than the numbers of weaned piglets being 0.8074 in 1992, 0.7145 in 1993, 0.7119 in 1994, 0.7433 in 1995, 0.4516 in 1996 and in 1997 only 0.4250

of the production volume achieved in 1991 (618,849 kg). Mean live weights of weaned piglets were 0.9919 in 1992, 0.9883 in 1993, 0.9756 in 1994, 0.9873 in 1995, 0.9242 in 1996 and in 1997 0.9254 of 6.059 kg registered for 1991. Just mentioned values resulted from the procedures applied in the farm where the durations of the occupation period of the farrowing rooms were 1.1507 in 1992, 1.2028 in 1993, 1.2789 in 1994, 1.1837 in 1995, 1.5478 in 1996 and in 1997 1.7778 of 28.79 days as it lasted in 1991. The numbers of grams of the piglets' live weight per day of the occupation period of the farrowing rooms were 181 in 1992, 173 in 1993, 161 in 1994, 175 in 1995, 125 in 1996 and in 1997 only 109 or 52,00% of 210 registered for 1991. Damaging of the piglets' ability to grow was the main reason of decreasing the production volume under unfavorable circumstances during the war and transitional years.

Keywords: Piglets, Growth

Averaging and diffusion approximation of the endemic SIR model in a semi-Markov random medium

Mariya Svishchuk

Mount Royal University, Canada

In the classical SIR epidemic model homogenous mixing has been assumed with all members of the host population equally likely to be in potentially infectious contacts. When the surrounding medium is heterogeneous (infectious contact rates are different for distinct subgroups of a population) it is natural to employ a dynamical model in random media and to look for an effective behavior through asymptotic analysis of the model based on the separation of scales.

We present an evolutionary system in a semi-Markov random medium as a modeling system. Through the process of averaging / homogenization we replace the complicated small scale model with an asymptotically equivalent homogeneous model. Diffusion approximation of the model in a semi-Markov random media makes it possible to investigate an asymptotic behavior of the solutions in the neighborhood of the endemic equilibrium. Under the additional balance condition we show that the initial system admits an approximation described by a system of stochastic differential equations.

Keywords: Evolutionary system, semi-Markov process, differential equation, endemic equilibrium, diffusion approximation

Delayed Heston Model: Improvement of Vol Surface and Hedging of Vol Swaps

Anatoliy Swishchuk

University of Calgary, Canada

We present a variance drift adjusted version of the Heston model which leads to significant improvement of the market volatility surface fitting

(compared to Heston). The numerical example we performed with recent market data shows a significant (44%) reduction of the average absolute calibration error (i.e., average of the absolute differences between market and model implied BS volatilities) (calibration on Sep. 30th 2011 for underlying EURUSD). The main idea behind our model is to take into account some past-dependent history (a.k.a delay) of the variance process in its (risk-neutral) diffusion. We also focus on volatility swap pricing and hedging.

Keywords: delayed Heston model; variance swap; volatility swap; hedging of volatility swap

Intergenerational mobility as a distance measure between probability distribution functions

M. Symeonaki, G. Stamatopoulou

Department of Social Policy, Panteion University of Social and Political Sciences, Greece

This paper presents a methodology of estimating social intergenerational mobility as a distance or similarity measure between the parent's probability distribution function and the sibling's. Several distance and similarity measures are provided and their properties are discussed. An illustration of the methodology is presented providing the measurement of the intergenerational occupational mobility with evidence drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE project), and more specifically from SHARELIFE release 1.0, concerning Greece¹.

Keywords: Intergenerational mobility, distance and similarity measures, SHARE project, occupational intergenerational mobility.

Asymptotic behavior of kernel density and mode estimates for censored and associated data

Tatachak Abdelkader¹, Ferrani Yacine², Ould Said Elias³

¹Lab. MSTD, Faculty of Mathematics, USTHB, Algeria, ²Faculty of Sciences, Department of Mathematics, UMBB, Algeria, ³Univ. Lille Nord de France, France

The aim of this paper is to investigate the asymptotic behavior of a kernel density estimator for randomly right-censored data under association hypothesis. As a result, we state the optimal uniform almost sure convergence rate of the estimator and then, as an application, we derive the almost sure convergence rate of a new kernel mode estimator for the true mode of the underlying density function.

Keywords: Associated data, Censored data, Kaplan-Meier estimator, Kernel mode estimation, Strong uniform consistency.

Minimum Pseudodistance Estimators and Applications to Portfolio Optimization

Aida Toma^{1,2}, Samuela Leoni-Aubin³

¹Department of Applied Mathematics, Bucharest Academy of Economic Studies, Romania, ²"Gh. Mihoc - C. Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania, ³INSA Lyon, France

The presence of outliers in financial asset returns is a frequently occurring phenomenon and may lead to unreliable mean-variance optimized portfolios. This fact is due to the unbounded influence that outliers can have on the mean returns and covariance estimators that are inputs in the optimization procedure. In the present paper we consider new robust estimators of location and covariance obtained by minimizing an empirical version of a pseudodistance between the assumed model and the true model underlying the data. We prove statistical properties of the new mean and covariance matrix estimators, such as affine equivariance, B-robustness and efficiency. These estimators can be easily used in place of the classical estimators, thereby providing robust optimized portfolios. A Monte Carlo simulation study and an application to real data show the advantages of the proposed approach.

Acknowledgements

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-RU-TE-2012-3-0007.

Modeling of correlated longitudinal data and estimation of seasonal factor of zoonotic cutaneous leishmaniasis cases in Center of Tunisia

Amine Toumi, Sadok Chlif, Jihene Bettaeib, Afif Ben Salah

Laboratory of Medical epidemiology, Pasteur Institute of Tunisia, Tunisia

Epidemiological data are often correlated and relationships between variables are not linear. Therefore, standard techniques of modeling and analysis are not applicable. Generalized additive (GAM) and generalized estimating equations (GEE) models seem useful to handle this kind of data. The estimation of seasonal factor, intra-annual and inter-annual cycles are important in epidemiological studies as it is considered as predictive tool and may pave the way for efficient early warning models of such disease emergence.

The seasonality of ZCL incidence was assessed by using a box diagram. The presence of significant seasonality was checked through three tests. Monthly stable seasonality test was done using Fisher (F_S) and Kruskal-Wallis tests. The moving seasonality test is considered as a two-factor ANOVA test (F_M). For these two tests, the significance level of Fisher test has to be set to 0.1% to ensure better precision of estimates. The combined test of presence of an identifiable seasonal variation aims to compare the relative importance of moving (F_M) and stable (F_S) as

$$T = \left(\frac{T_1 + T_2}{2} \right)^{1/2}$$

described in the following formula: , where T_1 is the ratio

$\frac{7}{F_S}$ and T_2 the ratio $\frac{3F_M}{F_S}$. If $T > 1$, the null hypothesis is rejected and we conclude that no identifiable seasonality is present as claimed by Lothian and Morry. X-12-ARIMA was performed on identifiable seasonality based on a series of moving average.

GAM and GEE models are extension of generalized linear models (GLM). The difference between GLM and GAM rise in the predictor which is linear in GLM and it is called additive in GAM. Spline functions, which are based on polynomial regression, were used in GAM models to identify if breakpoints existed. In GEE models, an autoregressive of order 1 for correlation matrix structure were used to obtain a valid statistical model.

The application of these techniques show the existence of inter-epidemic ZCL cycle ranging from 4 to 7 years and climate factors: rainfall above 37 mm and humidity had a significant effect on the incidence of the disease.

Keywords: GAM, GEE, X-12-ARIMA, seasonality, Leishmaniasis

Targeted minimum loss estimation of the causal effect of oldness of foci on ZCL incidence based on clustered data

Amine Toumi¹, Antoine Chambaz², Dhafer Malouche³, Afif Ben Salah¹

¹Laboratory of epidemiological department, Pasteur Institute of Tunis, Tunisia, ²Modal'X, Université Paris Ouest Nanterre, France,

³Engineering School of Statistic and Data Information

We evaluate the causal effect of oldness of foci on zoonotic cutaneous leishmaniasis (ZCL) incidence using targeted minimum loss estimation (TMLE, see van der Laan & Rubin, 2006, and Rose and van der Laan, 2011). Each subject contributes an observation $O = (C, V, W, A, Y)$ where the baseline covariate (C, V, W) includes potential confounders,

the exposure variable $A \in A$ is the (discrete) oldness of foci in the area where the subject lives, and the binary response variable Y indicates whether the subject is a ZCL case or not. Here, C stands for the description of the subject's household, $V \in V$ (a discrete set) for his/her age group, and W for his/her subject-specific covariates (such as sex, education...). The targeted parameter is $\psi : a \in A, v \in V$. It can be interpreted causally under standard causal assumptions.

The inference problem is not completely standard because the observations are clustered, households being sampled, not individual subjects (all subjects from the same household share a unique C and are dependent). To overcome this difficulty, we introduce a weighted version of the TMLE procedure, in the spirit of the case-control weighted TMLE procedure of Rose & van der Laan (2010). Under standard assumptions, the resulting targeted minimum loss estimator (TMLE) is double-robust (*i.e.*, consistent whenever one at least of two infinite-dimensional features of the data generating distribution is well estimated) and asymptotically gaussian.

We carry out a simulation study to illustrate the behavior of the TMLE and investigate its finite sample properties. By applying the methodology to the real dataset, we show that in old area and for subjects aged more than 25 years, they had a high risk to develop ZCL lesions while in new endemic are young subjects are more vulnerable to develop disease.

Keywords: Targeted estimator, Weighted, Clustered, Leishmaniasis

Stochastic variANts of the EM algorithm FOR PARAMETER ESTIMATION IN PLANT GROWTH MODELS

Trevezas Samis¹, Sonia Malefaki², Paul-Henry-Cournède¹

¹EPI DIGIPLANTE, INRIA, Laboratory of Applied Mathematics and Systems, Ecole Centrale Paris, France, ²Department of Engineering Science, University of Patras, Greece

The last decades advanced plant growth models have been proposed in the literature. Among these models, the GreenLab model (Cournède et al., 2006) has been proved to be very generic. In Cournède et al.[2] a parameter estimation approach was proposed for a large variety of plant and tree structures. Two assumptions of this approach concern an underlying deterministic model of biomass production and uncorrelated errors in the mass measurements of different types of organs in the plant structure. By relaxing these assumptions, in this study, we present a novel idea of modeling plant growth in the framework of non-homogeneous hidden Markov models (Trevezas and Cournède[3]) for a certain class of plants with known organogenesis (structural development). The model hidden states correspond to a sequence of unknown biomasses (masses measured for living organisms) produced during successive growth cycles and the observed variables correspond

to organ masses. For the parameter estimation, we compare two stochastic variants of a generalized Expectation-Maximization algorithm, one based on Markov Chain Monte Carlo (MCMC), see Trevezas et al.[4] and the other based on sequential importance sampling with resampling (SISR), see Trevezas and Cournède[3]), which correspond to two different ways of approximating the E-step. For the M-step, we use a conditional maximization approach. In this way, the parameters of the model are separated into two groups, one for which explicit updates can be derived by fixing the parameters of the other group, and one for which updates are derived via numerical maximization. The proposed methods are tested and compared on simulated and real data from the sugar-beet.

Keywords: plant growth model, non-homogeneous hidden Markov model, stochastic EM algorithm, MCMC, sequential importance sampling

References

- 1.Cournède P.-H., Kang M. Z., Mathieu A., Barczy, J.F., Yan, H.P., Hu, B.G. and De Reffye, P. Structural Factorization of Plants to Compute their Functional and Architectural Growth. *Simulation* **82**(7), 427-438, 2006.
- 2.Cournède P.-H., Letort, V., Mathieu, A., Kang, M.-Z., Lemaire, S., Trevezas, S., Houllier, F. and de Reffye, P. Some parameter estimation issues in functional-structural plant modelling, *Mathematical Modelling of Natural Phenomena*, **6**, 133–159, 2011.
- 3.Trevezas, S. and Cournède P.H. A sequential Monte Carlo approach for MLE in a plant growth model, (*submitted*).
- 4.Trevezas S., Malefaki S. and Cournède P. H. Simulation techniques for parameter estimation via a stochastic ECM algorithm with applications to plant growth modeling, (*submitted*).

FUNCTIONAL MODELS TO FORECAST AIRBORNE POLLEN CONCENTRATION AS PREVENTIVE TOOL FOR ALLERGIES

Mariano J. Valderrama, Manuel Escabias

Department of Statistics and Operations Research, University of Granada, Facultad de Farmacia, Spain

The health problem of allergy caused by airborne pollen concentration is increasing more and more along the years especially in some geographic areas as Mediterranean region. Of course there exists a different cause of allergy due to mites or other animals, but this is much more studied and its effects controlled. So, forecasting pollen levels is an interesting problem not only from an environmental point of view but also in health-care planning, mainly in vaccination strategies related to allergies among children and the elderly. Furthermore, an economic impact must be taking into account because of the sick leaves and also residence removal it causes. Perhaps cypress and olive trees have the highest pollen incidence and are present in the atmosphere practically all year round, although they are predominant at the end of the winter and

spring period, making this particle a powerful allergen. In Europe, allergy to Cupressaceae pollen was considered a rarity until 1975, but is now a recognized clinical entity.

In order to develop a stochastic model to explain this phenomenon, different methodologies have been applied considering also several meteorological covariates such as temperature, isolation, humidity and rainfall on the previous days. The influence of the wind varies from a zone to another without a well-defined pattern of association. According to temperature it can have influence in different ways, for example it is different the influence of maximum, minimum and mean temperatures, and it is not in the same way in all ranges, we can find less influence behind a temperature level and greater over it. The range of temperatures can also influence pollen concentration. Moreover the influence is not the same everywhere, different cities can have different factors that have influence in airborne pollen concentration.

There are two main ways for prediction in time series. The first includes all the self-explicative and consider only the past history of the series itself. The second one is based on dynamic regression and includes information from an input process, with the inertia process also being represented as an ARIMA model. Nevertheless neither of the above techniques are suitable for forecasting peak pollen concentrations due to the very sparse nature of the data and especially because there are discontinuities in the atmospheric presence of pollen.

The most novelty and effective methodology to deal with this problem is the functional data approach that considers curves as data. At the same time, the functional logistic model estimates a binary response variable from a functional predictor and it is suitable to solve the problem of forecasting the pollen concentration in terms of a set of explanatory variables.

Keywords: Functional data, principal component, ARIMA, airborne pollen

A Dairy Market Information and intelligence System

Arturo Valdivia¹, José A. Espinosa², Arturo G. Valdivia³, Georgel Moctezuma², Alejandra Vélez², José L. Jolalpa², Arturo González², América A. Luna², José J. Espinoza², Sergio F. Góngora², Carlos M. Arriaga-Jordán² y Francisco E. Martínez⁴, José L. Dávalos⁵

¹University de Barcelona, Spain, ²National Institute for Agricultural, Livestock and Forestry Research of Mexico, Mexico, ³Autonomous University of Aguascalientes, Mexico, ⁴Autonomous University of Mexico State, Mexico, ⁵National Autonomous University of Mexico, Mexico

In this paper we present the implementation and initial results of the project InfoLeche-México, an information and intelligence system for the bovine milk market in Mexico. The aim of InfoLeche-México is twofold. Firstly, to measure and collect up-to-date relevant data related to specific

milk market variables and production costs of dairy farms all across the Mexican territory. Secondly, to analyze the collected data in order to model and compute a series of production indicators that may help milk producers to take better decisions and consequently improve their business.

The variables of interest were selected following the Delphi methodology, including farmers and experts from several Mexican research institutions. As a result of this panel of experts, nearly 1450 variables were defined. The up-to-date value of these variables is uploaded to a content management system which acts both as data warehouse and knowledge community environment. Once the data is the system, a series of target variables are computed and interactively presented to the user. All the processed data is public but previously anonymized. In addition, every month the research group of the project elaborates a newsletter summing up a series recommendations.

Keywords: Computing-aided decision support, Data warehouse, Delphi methodology, Stochastic modelling, Risk indicators, Time series analysis, Knowledge systems, open-source content management system

**Equity in social policy: a human rights and results based approach - A welfare state comparison: the case of disability
Working paper, September 2012**

J Van Geystelen, E P Poirot

The reinforcement of equity and the inclusion of vulnerable population groups in social welfare systems development remains prominent in both, local and global policy arena. This paper assesses the hypothesis that certain welfare regimes produce more equitable social welfare outcomes than others. A human rights and results based perspective is adopted to explore whether there may be a relation between social welfare outcomes in the general population and in less advantaged population groups across welfare systems with different classification. Welfare sample states are consistent with a tripartite classification of welfare regimes, namely: a) social: Denmark, Norway, Sweden; b) corporatist: France, Germany; and c) liberal: United Kingdom, United States of America. For the purpose of this paper "people living with disability" represent vulnerable population groups. A choice founded on the universal characteristics of disability, including its complexity that requires a holistic approach with strong cross sector collaboration. Results from both, the general and vulnerable population were tabulated to calculate a human rights and results based equity index (HREI). The integration of an additional cost factor allowed to explore cost-

effectiveness. Findings confirm the hypothesis that certain welfare regimes do produce more equitable social welfare outcomes than others with the social regime to produce best and most equitable results as well as best appreciation of cost-effectiveness. The latter could indicate that welfare systems founded on limited de-commodification, high public investment, prioritisation of employment and valuing of collectivity and solidarity are most efficient. Yet, more in-depth research would be required to validate these findings and to better understand causality.

Keywords: Equity, social policy, welfare state comparison, results based approach, disability.

A full version of this paper can be requested at: social.12d808@gmail.com

Assessing errors in CMM measurements via Kriging and variograms: a simulation study

Grazia Vicario, Suela Ruffa, Giovanni Pistone

Politecnico di Torino, Italy

Industrial parts are routinely affected by dimensional and geometric errors due to the manufacturing processes used for their production.

These errors, that usually have a typical pattern related to the employed manufacturing process, are limited by means of dimensional and geometrical tolerances (such as straightness, roundness, flatness, profile) that have to be verified on the manufactured parts. Coordinate Measuring Machines (CMM) are the most common equipment for 3D measurement because of their accuracy and flexibility.

In the present paper we focus on the inference on the error of different planar surfaces whose tolerances are verified using a CMM. For this purpose we suggest the prediction of the surface model using a Kriging model on a set of measured points. This method was developed by George Matheron as the "theory of regionalized variables" and by D.G. Krige, who named it, as an optimal method of interpolation for use in the mining industry.

Basically, Kriging is a linear predictor of unknown values from data observed at known points (training data) and the predicted surface passes through all the known points. The rationale behind spatial interpolation is the observation that points close together in space are more likely to have similar values than points far apart; that's way the 'neighboring' observations gain more weight. Technically, Kriging is a

stochastic linear interpolation technique that predicts the response values at untried locations with weights assigned to the tried locations. The weights are selected so that the estimates are unbiased (repeatedly using Kriging, we expect the correct result on average) and they have minimum variance.

Therefore, the fundamentals is the rate at which the variance between points changes over space. This is expressed as a variogram which shows how the average difference between values at points changes with distance between points; it is a function of the distance and of the corresponding direction of any pair of points depicting their correlation extent. Theoretically, it is defined as the variance of the difference between the response values at two locations and it is equivalent to the correlation function for stationary processes. The use of the variogram instead of the correlation function is recommended by the geostatisticians even if the process is not stationary, because of possible contaminations by trend estimate residuals.

In studying the variogram it is necessary to make some specifications about the origin of the observed variation on the surface.

Consequently, it is essential to resort to a proper model: Simple Kriging, i.e. assuming that the surface has a constant mean, no underlying trend and that all variation is statistical and Universal Kriging, i.e. assuming that there is a deterministic trend in the surface that underlies the statistical variation. In both the cases, once trends are assumed not to exist or they have been accounted for, any other variation is hypothesized to be a function of distance.

Since the trend mostly reveals the global characteristics of the phenomenon by means of the evaluation on large scale variations, the random component of the response values may be enlightened by the variogram, that accounts local irregularities (small scale variation).

In this paper we resort to variograms to detect possible manufacturing signatures of the production process and systematic errors of the CMM measurement process. We simulated different, and most common, manufacturing signatures of a planar surface and possible errors of a measurement process with CMM. Then, the variograms have been estimated, using the most robust empirical estimator in the case at hand and the likelihood (or restricted likelihood) estimator. The behavior of the omnidirectional variogram suggests the spatial correlations, giving evidence of possible non isotropy.

Keywords: Kriging, Variogram, Coordinate Measuring Machines

Estimating the model with fixed and random effects by a robust method

Jan Ámos Víšek

Institute of Economic Studies, Faculty of Social Sciences, Charles University, Czech Republic

Regression model with fixed and random effects estimated by modified versions of the *Ordinary Least Squares* (OLS) is a standard tool of panel data analysis. However, it is vulnerable to the bad effects of influential observations (contamination or atypical observations). Robustification of the OLS by means of the *Least Weighted Squares* (LWS) and modifications of LWS in an analogous way as the OLS were modified for estimating the fixed- and random-effects-model rid of this disadvantage. The numerical study reveals the reliability of the respective algorithm. The results of this study were collected in a file which is possible to find on web, address is given below. Patterns of these results were included into the paper (extent of which is limited). The possibility to reach nearly the full efficiency of estimation in the case when there are no influential points is demonstrated at the mentioned file on the web.

Keywords: Linear regression model, the least weighted squares, fixed and random effects, numerical simulations.

Estimating the model with fixed and random effects by a robust method

Jan Ámos Víšek

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague

A typical observation in a data set can cause misleading conclusions of the regression analysis. That was the reason for building up the robust methods for identifying the true underlying model in the various frameworks. One such framework is the regression model with fixed and random effects and we can meet with proposals of estimators based on the idea of *M-estimators*, of the *Least Median of Squares* or of the *Least Trimmed Squares*, see e. g. [2], [3], [5], [6], [7] or [8]. The former require studentization of residuals – to reach *scale-* and *regression-equivariance* of the respective estimators, the latter can be very sensitive to „inliers“, see [9] or [10], compare also [4]. The *Least Weighted Squares* which employs the idea of smooth decrease of the influence of atypical

observations by means of prescribing the weights to the order statistics of the squared residuals rather than to the squared residuals directly, rid of both these problems. The advantage of the method is that it can be adjusted to the level of contamination by an adaptive selection of the weights. Its generalization for the model with fixed and random effects is studied and an extensive numerical simulations (part of which will be presented) exhibit its behaviour for the finite samples.

(The research was supported by the Czech Science Foundation project No. 13-01930S.)

Keywords: The linear regression model, least weighted squares, fixed and random effects, numerical simulations.

References

- [1] Bickel, P. J. (1975): One-step Huber estimates in the linear model. *JASA* 70, 428--433.
- [2] Bramati, M. C., Croux, C. (2007): Robust estimators for the fixed effects panel data model. *Econometric Journal* 10, 521 - 540.
- [3] Dehon, C., Gassner, M., Veradi, V. (2009): Beware of good outliers and overoptimistic conclusions. *Oxford Bulletin of Economics and Statistics* 71 (3), 437 - 452.
- [4] Hettmansperger, T. P., S. J. Sheather (1992): A Cautionary Note on the Method of Least Median Squares. *The American Statistician* 46, 79--83.
- [5] Kott, P. (1989): Robust small domain estimation using random effects modelling. *Survey Methodology* 15, 1 - 12.
- [6] Rocke, D. M. (1991): Robustness and balance in the mixed models. *Biometrics* 47, 303 - 309.
- [7] Veradi, V., Croux, C. (2009): Robust regression in Stata. *The Stata Journal* 9 (3), 439 - 453.
- [8] Veradi, V., Wagner, J. (2010): Robust estimation of of linear fixed effects panel data model with an application to the exporter productivity premium. *Discussion paper No 4928, IZA, Bonn*.
- [9] Víšek, J. Á. (1996): A cautionary note on the method of Least Median of Squares reconsidered. *Trans. Twelfth Prague Conf. on Information Theory, Statistical Decision Functions and Random Processes, Academy of Sciences of the Czech Republic*, 254 - 259.
- [10] Víšek, J. Á. (2011): Consistency of the least weighted squares under heteroscedasticity. *Kybernetika* 47, 179-206.

MULTIVARIATE LIFETIME MODELS BASED ON LATENT RANDOM VARIABLES: A REVIEW AND SOME RECENT EXTENSIONS*

Juana-María Vivo

University of Murcia, Regional Campus of International Excellence
"Campus Mare Nostrum", Murcia, Spain

Multivariate lifetime data frequently arise in many fields such as medicine, biology, public health, epidemiology, engineering, economic and demography. Undoubtedly, it is important to consider different multivariate distributions that could be used to model such multivariate lifetime data, and their properties are also useful to carry out that purpose. In the univariate case, exponential, Weibull and gamma distributions have been some of the most frequently applied statistical distributions in reliability and survival analysis, and multivariate extensions have been constructed from these models. However, the multidimensional extension of a distribution model is not unique, i.e., different derivations through the construction methods could be called multivariate extensions.

In this work, we focus on the construction of the multivariate models based on the method of the latent random variables which was used by Marshall and Olkin in 1967 to introduce a multivariate exponential distribution. A brief review about this technique is shown along with some recent extensions and their properties.

Keywords: Latent factor, Exponential distribution, Multivariate distributions, Failure rate

References

[1] Marshall, A.W., Olkin, I., A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44, 1967.

* This work was partially supported by Fundación Séneca of the Regional Government of Murcia (Spain) under Grant 11886/PHCS/09.

Kernel Density Derivatives Estimations with Iterative Bandwidth Selection

Kamila Vopatová¹, Ivanka Horová², Jan Kolářček²

¹Department of Econometrics, University of Defence, ²Department of Mathematics and Statistics, Masaryk University, Czech Republic

Kernel estimates belong to a class of effective nonparametric estimations of a probability density and its derivatives. Significant information about features of the density is contained in the first and the

second derivative of the true density. These estimates depend on a bandwidth parameter which controls the smoothness of the estimates.

A special iterative method for multivariate kernel density derivative estimators is presented. The advantage of the proposed method consists in the fact that it does not need any pre-transformation of the data or an additional bandwidth parameter as in plug-in methods.

The proposed method is compared with the plug-in method and the least square cross-validation method through a simulation study and real data sets as well.

Keywords: Density estimate, iterative method, kernel smoothing.

CLASSIFICATION ALGORITHMS AND FEATURE SELECTION SCHEMES FOR PREDICTING STUDENTS' SUCCESS - A CASE STUDY

Milan Vukićević, Miloš Jovanović, Boris Delibašić, Milija Suknović

Faculty of Organizational Sciences, University of Belgrade, Serbia

Educational Data Mining is a new emerging field that discovers knowledge from data originating from educational environments. It has a focus on different aspects of educational objects (e.g. students, teachers, teaching materials, organization of classes etc.). One of the most frequent research topics in the area of EDM (also investigated in this research) is the prediction of student performance. The main idea behind this research direction is that based on student activity one can predict the future outcome of student performance. The main use is early analysis of students' success on each study program and recommendation of a study program for which a student will likely have good success in. In this research we tried to identify the best method for selection of features for different classification algorithms, for prediction of students' success. Several advanced feature selection schemes are

evaluated on! how effective they are for each classification algorithm. Optimal combination of features and classification algorithms are used for early analysis of students' success on each study program and recommendation of a study program. For the purpose of data analysis and model evaluation, in this work we used RapidMiner, a widely used open source software solution for data mining, which provides large number of tools for tackling the overall data mining process.

Keywords: Educational Data Mining, Feature selection, Classification, RapidMiner

HOME VERSUS HOSPITAL CANCER MORTALITY (2006-2011) AND PALLIATIVE CARE

Elena Vvedenskaia

Medical University, Russia

Objectives. Our aim was to examine the proportion of home versus hospital chronic progressive diseases mortality in a large Russian region in relation to demographic factors and to identify 5 year trends in home deaths from cancer. **Materials and methods.** The data on the place of death were derived from death registrations for all cancer deaths in the region between the years of 2006-2011. We examined 5 year trend in home and hospital mortality for the whole region, for the city and for the rural area. **Results.** Within this period the percentage of home deaths from all causes significantly increased: in the region from 55.67% to 73.68%, in rural areas from 60.10 to 82.12% and in the city from 30.32% to 66.37%. The number of deaths from cancer was consistently high and the rate of home cancer deaths was 85.0-89.0% (87.52% on average) for these years. Older people with cancer were more likely to die at home (85.16% of all cancer deaths). Among all women and men died from cancer respectively 90.62% and 79.60% died at home. The proportion of home and hospital deaths from cancer was 87.52 and 7.48%, respectively. **Conclusions.** High proportion of home deaths is the

result of reducing the hospital care supply and demonstrates low availability of inpatient palliative care. These findings today highlight the need for a palliative care service establishment for patients with life-limiting and life-threatening diseases and special attention should be paid to home care services establishment and palliative care education of primarily care medical specialists.

Keywords: place of death; home cancer deaths, hospices; palliative care; home care services

Frequentist Model Averaging for Multinomial and Ordered Logit Models

Alan Wan, Xinyu Zhang

City University of Hong Kong, Hong Kong

Multinomial and ordered Logit models are quantitative techniques used in many disciplines nowadays. When applying these techniques, practitioners usually select a single model by information-based criteria or pretesting. In this paper, we consider the alternative strategy of combining models instead of selecting a single model. Our strategy of weight choice for the candidate models is based on the minimization of a plug-in estimator of the asymptotic squared error risk of the model average estimator. Theoretical justifications of this model averaging strategy are provided, and a Monte Carlo study shows that forecasts produced by the proposed model averaging scheme are often more accurate than forecasts based on some of the common model selection schemes. An empirical example based on credit rating data is used to illustrate the proposed method. To reduce computational burden we also consider a model screening strategy that eliminates some of the very poor models before averaging.

Keywords: asymptotic squared error risk; local mis-specification; model screening; Monte Carlo; plug-in estimator

A Variable Window Scan Statistic for MA(1) Process

Xiao Wang, Joseph Glaz

Department of Statistics, University of Connecticut, USA

Approximations for the distribution of a scan statistic for independent and identically distributed (iid) observations from a normal distribution have been investigated in Glaz et al. [5]. Based on these approximations, a variable window scan statistic has been introduced in Wang and Glaz [6]. In this article we derive a product-type approximation and Bonferroni-type inequalities for the distribution of a scan statistic for observations from a moving average process of order one, where the error component has a normal distribution with mean of zero and variance of one. We also investigate the performance of an approximation for a sequence of moving sums of 1-dependent observations that has been derived in Haiman [7,8]. A nice feature of this approximation is that an error bound can be evaluated as well. Numerical results to evaluate the accuracy of the approximations and inequalities are obtained via an algorithm established by Genz and Bretz [9]. We also present an algorithm to implement a variable window scan statistic.

Keywords: Moving Sum, Approximation and Inequality, Moving Average, R Algorithm.

A hybrid Monte Carlo method for high-dimensional option valuation

Antony Ware, Binbin Wang

University of Calgary, Canada

We propose a hybrid Monte Carlo method for high dimensional option valuation. The approach can be viewed as an approximate conditional Monte Carlo method. The option value is written as a nested expectation by isolating a lower dimensional (usually one-dimensional) random variable, and the inner expectation with respect to this variable is approximated using a Fourier method (based on an expansion in terms of sine and/or cosine functions). It is shown that the bias introduced by

this approximation can be controlled using a multiscale approach without compromising the efficiency of the estimator. Applications to basket options, Asian options are demonstrated.

Keywords: Monte Carlo option pricing Fourier methods

A fuzzy correlation coefficient for bivariate fuzzy data

Erika Watanabe¹, Kyohei Yamamoto¹, Norio Watanabe²

¹Graduate School of Chuo University, ²Department of Industrial and Systems Engineerings, Chuo University, Japan

Personal feeling is subjective and can be represented by a fuzzy set in some circumstances. In this study we assume that bivariate data are given by fuzzy sets and consider the correlation analysis of such fuzzy data. Correlation coefficients for fuzzy data have been proposed by some authors already. Some coefficients are crisp numbers and some are fuzzy numbers. However, it is not clear which coefficient we should use. In this study we propose a new fuzzy correlation coefficient from the practical viewpoint. Usability of the proposed correlation coefficient is illustrated by applying to real fuzzy data obtained by a questionnaire.

Keywords: fuzzy random variable, kansei data

Decomposition of trends by a fuzzy trend model for multivariate time series

Norio Watanabe, Yasutaka Ohta

Chuo University, JAPAN

Analysis on trend is one of the most important issues in time series analysis. For multivariate time series trends of each series can be decomposed into common and individual trends usually. The multivariate fuzzy trend model has been proposed for such decomposition of trends. In this study we assume that time series can be classified into some groups and propose an extended fuzzy trend model which can decompose trends into common trend, group trends and individual trends. An identification method is provided by using three-stage least square method and an information criterion. We also propose indices for measuring weights of each trends. Usability of the proposed model is demonstrated by applying to real time series.

Keywords: common trend moving average

Optimal Designs for Heteroscedastic Accelerated Life Testing Models with Two Factors

Xiaojian Xu, Mark Krzeminski

Brock University, Canada

For any highly reliable product, it takes too long to run to yield desired information on its life under normal use condition. Accelerated life testing (ALT) provides a means of shortening its life and quickly obtaining data by subjecting the product to higher-than-normal-use levels of the stress factors. In this paper, we discuss the construction of the optimal designs for two-factor accelerated life tests in order to estimate percentiles of product's life distribution under normal use condition. We assume a Weibull life distribution, log-linear life-stress relationships with heteroscedasticity (non-constant shape parameter) for both stress factors. There is no interaction between two factors. Time-censoring is utilized for the experiment termination. The primary optimality criterion is the minimization of the asymptotic variance of MLE of the percentile of interest at normal use stress level. This yields infinitely-many such designs, therefore the optimal one is further chosen by a secondary criterion, which is to maximize either the determinant or the trace of Fisher information matrix. General methods of finding the optimal designs are discussed. The resulting designs are illustrated with some practical examples from engineering and a comparison study is also presented, which compares the results between the two secondary optimality criteria and also the results between the cases of homoscedasticity and heteroscedasticity.

Keywords: Time-censoring, shape-stress relationship, log-linear model, equal probability curve, A-optimality, D-optimality, degenerate designs, non-degenerated designs.

Guidelines to Assess a Weibull Approximation to an Independent Competing Risks Model with Two Weibull Distributions

Sergio Yáñez¹, Luis A. Escobar², Nelfi González¹

¹Escuela de Estadística, Universidad Nacional de Colombia,

²Department of Experimental Statistics, Louisiana State University, USA,

³Escuela de Estadística, Universidad Nacional de Colombia, Colombia

In different application areas, e.g., medicine, biological sciences, and engineering reliability, competing risks models are used to model life of systems with multiple failure modes. The observed failure is the minimum of the possible individual failure times. We discuss the distribution of the minimum for a competing risks model with two

independent failure modes and Weibull distributed. We develop criteria and guidelines to assess a Weibull approximation to the distribution of the minimum. We present a relation between a Kullback-Leibler procedure to obtain the best Weibull approximation to the distribution of the minimum and the maximum likelihood Weibull fit ignoring the mode of failure information. The study considers this relationship in small and large samples.

Keywords: Competing risks, Kullback-Leibler information, Ignoring mode of failure, Weibull distribution, Multiple failure modes, Maximum likelihood.

This work has been supported by Departamento Administrativo de Ciencia, Tecnología e Innovación-Colciencias, República de Colombia. Código 110152128913.

The Algorithm for the Faultless Work Probability Determination of the Water Pipeline Networks at the External Influences

Boli Yarkulov

Samarkand State Institute of Building and Architecture, Uzbekistan

The analysis of the managing enterprise system of water-supply shows that its administrative body has to control the managing object based on its condition. In order to organize the control of the managing object based on its condition, the following concept is necessary:

- 1) The determination of the reliability characteristics that shows the managing object's actual condition.
- 2) The determination of the main parameters of the controlling (spare parts, workers, and employees) technical maintenance on the basis of reliability characteristics of control object.

One of the steps of the concept in the above is to determine the probability of faultless work of water pipeline networks at external influences i.e. earthquake.

Keywords. Water pipeline networks, reliability, algorithm, failures, probability.

Asymptotic Behavior of Green's Functions Arising in Branching Random Walks

Elena Yarovaya

Department of Probability Theory, Moscow State University, Russia

Information about asymptotic behavior of the transition probabilities is of great importance in analysis of properties of continuous-time branching

random walks on multidimensional lattices. The behavior of transition probabilities of a branching random walk in the situation when the space and time variables grow jointly is established. Main results here are the limit theorems about properties of the Green function for the transition probabilities. These results are applied for the investigation of the large deviations for branching random walks, in particular, for studying of the particle population front.

Keywords: Large Deviations, Green's Functions, Branching Random Walks.

Multifractal analysis: identifying the boundaries application in the study of HRV

Artem Zagaynov, Valery Antonov, Vu van Quang Saint-Petersburg

State Polytechnic University, Russia

The report focuses on the search for the control parameters of the computational procedure of the method wavelet transform modulus maxima to establish the most accurate estimates of multifractal component of HRV. The above-mentioned parameters can be attributed as the various properties of the wavelet basis function (and hence the behavior of the scaling exponent in different parts of the spectrum) and artificial introduction of the frequency characteristics of its expression (for correlation with wave processes that characterize the original time series). We have obtained the first results for HRV in health and in various disease (heart failure, atrial fibrillation, etc.). Based on these results suggested a system of monitoring HRV multifractal with automatic selection of the wavelet function generator capable of classification of certain kinds of extrasystoles in different parts of the spectrum and identify cardiovascular pathology developed methods of making automatic frequency parameters in the expression constructed function.

Keywords: wavelet transform modulus maxima

The reported study was supported by RFBR, research project No. 12-08-31108-mol_a

[1] V. Antonov, J. Danilevich A. Kovalenko, A. Zagajnov Dynamic-entropy monitoring daily state human heart rates // Abstracts The XIII International Conference «Applied Stochastic Models and Data Analysis» (ASMDA-2009). Vilnius 2009. Vilnius «Technika» P. 8.

[2] V. Antonov, A. Kovalenko, A. Zagaynov, Vu van Quang The research of fractal characteristics electrocardiogram in real time // Book of Abstracts International Conference «Stochastic Modeling Techniques and Data Analysis» (SMTDA-2010). Chania Crete Grece 2010. MAICH Prefecture of Chania, P.11

Statistical Inference for Partially Observed Linear Systems

Vladimir Zaiats

Department of Digital Technologies and Information, Escola Politècnica Superior, Universitat de Vic, c./ Laura, 13, 08500 Vic (Barcelona), Spain

We would like to focus on models describing the so-called partially observed linear systems, where the function we would like to control is not observed directly. The mathematical description of the model involves different functions, and our aim is to obtain estimates of these functions.

Assume that we observe a process $X = (X_t, 0 \leq t \leq T)$ satisfying the following system of stochastic differential equations:

$$\begin{aligned} dX_t &= h_t Y_t dt + \varepsilon dW_t, & X_0 &= 0, \\ dY_t &= g_t Y_t dt + \varepsilon dV_t, & Y_0 &= y_0 \neq 0, & 0 \leq t \leq T, \end{aligned}$$

where V_t and W_t , $0 \leq t \leq T$, are independent Wiener processes. The process $Y = (Y_t, 0 \leq t \leq T)$ is not observed directly, but this is the process which should be controlled. In this model, we look for statistical estimators of the functions $f_t = h_t y_t g_t$, h_t , y_t , $0 \leq t \leq T$, when the noise is small, that is, when $\varepsilon \rightarrow 0$. We obtain estimators which have optimal rates of convergence in the minimax sense.

This is a joint work with Yu. Kutoyants (Université du Maine, Le Mans France). The author is partially supported by the Ministerio de Economía y Competitividad under grant MTM2012-31118.

Keywords: partially observed linear systems, functional estimates, optimality, minimax.

Integrated Likelihood in Statistical Inference

Aleksander Zaigrajew

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Poland

Let us consider statistical models with two parameters: a parameter of interest and a nuisance parameter. We concentrate only on situations when the nuisance parameter is either location or scale. Assume that our goal is to estimate the parameter of interest or to test a simple hypothesis about it. The well-known approach to these problems is based on the maximum likelihood principle. For point estimation we maximize the likelihood function (or, to be more precise, the so-called profile likelihood function), while for confidence interval estimation or for testing hypotheses we base our inference on the likelihood ratio statistic. In the talk we propose an alternative approach to these problems. This

approach is based on making use of the integrated likelihood function. The integrated likelihood function is the integral of the likelihood function, with a weight function, over the nuisance parameter. Under some regularity conditions on distributions, we investigate the properties of the maximum integrated likelihood estimators and compare them with those of the standard maximum likelihood estimators. Also we study an effect of substituting the standard likelihood ratio statistic with that based on the integrated likelihood. In the area of testing hypotheses we also consider the problem of choosing between two families of distributions with the same support and compare the standard approach based on the ratio of maxima of two standard likelihood functions with that based on the ratio of maxima of two integrated likelihood functions. As examples of the results obtained different two-parametric families of distributions are considered. The results are supported with numerical calculations.

Keywords: Maximum likelihood principle, Integrated likelihood function, Likelihood ratio statistic.

ON THE PERFORMANCE OF ARTIFICIAL NEURAL NETWORKING FOR PROCESS MONITORING

Babar Zaman^{1,2}, Muhammad Riaz³

¹Department of Statistics, Quaid-Azam-University Islamabad, Pakistan,

²Research Department of King Khalid Eye Specialist Hospital, KingDom of Saudi Arabia, ³Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Statistical Process Control Charts (SPCC) is very powerful tools to monitor and ensure stability of the parameters of a process whether manufacturing or non-manufacturing. There are mainly three types of control charts namely Shewhart, EWMA and CUSUM. The major limitation of the Shewhart type control charts is that these are not very efficient in detecting small process shifts. To increase the sensitivity of Shewhart control charts for small shifts additional supplementary runs rules have been suggested in literature but they suffer with the inflation of false alarm rate. To address particularly the smaller shifts EWMA and CUSUM type charts serve the purpose efficiently.

Artificial Neural Network (ANN) technique is also nowadays popular to be used for monitoring process parameters as an alternative to SPCC due to its application and superior performance. This research focuses on the ANN technique for monitoring process spread for small, moderate and large shifts. The performance of the proposed ANN scheme is evaluated in terms of Average Run Length (ARL) and is also compared to SPCC for spread parameter. The ARL study has shown that the proposed ANN for spread is more efficient than the traditional control charts namely Shewhart's R and EWMA charts for both small and large shifts. We have also examined the robustness of the proposed ANN scheme.

Moreover, we have suggested the use of runs rules with ANN structure in the form of 2/2, 2/3 and 3/3 schemes. We have observed that the application of these runs rules schemes with ANN have the ability to perform good for different types of shifts as compare to the usual namely 1/1 scheme.

KEYWORDS: Artificial Neural Network (ANN); Average Run Length (ARL); Control Charts; Process Spread; Runs Rules; Statistical Process Control (SPC).

On the Choice of Runs Rules for efficient Process Monitoring

Babar Zaman¹, Muhammad Riaz^{2,3}

¹Research Department, King Khalid Eye Specialist Hospital, KSA,

²Department of Mathematics and Statistics, King Fahad University of Petroleum Minerals, Saudi Arabia, ³Department of Statistics, Quaid-i-Azam University, Pakistan

Statistical Process Control (SPC) charts are used to monitor the performance of a process by differentiating natural and unnatural variations. The Shewhart type charts are meant to address larger process shifts in the parameters of the quality characteristic of interest. These charts are not very efficient at detecting shifts of smaller magnitude, for which implementation of the runs rules schemes is an attractive option. In this study we have suggested the use of runs rules schemes for \bar{X} , R and S charts for Burr distributed process environments with different amounts of skewness and kurtosis (normal distribution is also considered as a special case). We have also investigated the choice of an appropriate runs rule for a given situation to monitor location and dispersion parameters. For performance evaluations and comparisons we have used probability to signal as performance measure. We have also given an application with a real dataset to elaborate and highlight the practical importance of the proposals of the study.

Keywords: Average Run Length (ARL); Burr Distribution; Runs Rules Schemes, Shewhart Charts, Skewness and Kurtosis.

Regression Trees, Boosted Regression Trees, and Markov Chain Monte Carlo based Imputation for the Prediction of the Strength Properties of Composite Materials

Yan Zeng¹, Dillon Carty², Timothy M. Young², Frank M. Guess¹

¹Department of Statistics, Operations, and Management Science, University of Tennessee, ²Center for Renewable Carbon, University of Tennessee, USA

Regression tree, boosted regression trees, and imputation with Markov Chain Monte Carlo (MCMC) simulation methods were employed to

predict the strength properties and improve the understanding of process relationships for wood composites manufacture. The data set was developed in 2009 from an industrial particleboard manufacturer located in the southeastern United States. The response variables were modulus of rupture (MOR) strength and tensile strength internal bond (IB). The complete data set consisted of 4,437 destructive laboratory tests spanning 15 months which was fused with 189 independent process variables from the production line data warehouse. The data set without imputation had 4,307 records, 130 records had incomplete data and were deleted for regression tree and boosted regression tree modeling. Eighty percent of the data were used for training and 20% were used for validation. The root mean square error of prediction as a percent of the mean (RMSEP%) for regression trees was 14.2% for IB tensile strength and 10.7% for MOR in the validation data sets. Boosted regression trees using stochastic gradient boosting had RMSEP% that converged at 12.7% for IB tensile strength and 8.5% for MOR. The correlation between the observed values and the predicted values was 0.86 for IB and 0.91 for MOR. The set after imputation with Markov Chain Monte Carlo (MCMC) consisted of the original 4,437 records. Eighty percent of the complete data after imputation were used for training and 20% were used for validation. The RMSEP% with MCMC imputation converged to 9.1% for IB tensile strength and 6.5% for MOR. The correlation between the observed values and the predicted values was 0.90 for IB and 0.94 for MOR.

Strong Invariance Principle (SIP) for Risk Processes with Stochastic Premiums

Nadiia Zinchenko

Department of Applied Mathematics, Faculty of Physics and Mathematics, Nizhyn State Mukola Gogol University, Ukraine

The Cramer-Lundberg risk model with stochastic premiums, which is a natural generalization of the classical risk model, is considered. Within such model both total claim amount process and total premium amount process are compound Poisson processes. We used various methods for investigation ruin probabilities in such models: solution of integro-differential equations, martingale technique, empirical de Vylder approximation and diffusion approximation based on weak convergence of risk process. But mainly we focused on the strong approximation of the risk processes. Our main tool is rather general SIP-type theorems for the superposition of random processes $S(N(t))$, when càd-làg random processes $S(t)$ and $N(t)$ themselves admit strong approximation by a Wiener or stable Lévy process. Such results serve as a source of numerous strong limit theorems for random sums under various assumptions on the summands and counting renewal process $N(t)$. As a

consequence a number of SIP-type theorems for risk processes in classical and renewal risk models are obtained as well as for risk processes with stochastic premiums and used for investigation the rate of growth and fluctuations of mentioned processes. Various modifications of the LIL and Erdős-Renyi-type SLLN are proved both for the cases of small and large claims.

Keywords: risk process, stochastic premiums, random sums, strong limit theorems, strong invariance principle.

Semi-parametric models for coral reef dynamics

Kamila Zychaluk¹, Matthew Spencer², Damian Clancy¹, John F. Bruno³ and Tim McClanahan⁴

¹ Department of Mathematical Sciences, University of Liverpool, UK

² School of Environmental Sciences, University of Liverpool, UK

³ Department of Biology, University of North Carolina, Chapel Hill, USA

⁴ Wildlife Conservation Society, Marine Programs, New York, USA

There are many mathematical models for the dynamics of coral reefs. Typically, these models assume the functional relationships that are responsible for changes in the reef community but there is often little evidence on which to choose the functional relationships. Furthermore, the parameters of such models are difficult to estimate. Instead, we propose a statistical model based on many data but relatively few assumptions. We use a large database of repeated observations of the composition of coral communities to make predictions about the dynamics of reef composition. We use our model to estimate a regional dynamic equilibrium in reef composition.

We have observations of the proportion of space occupied by three components (hard corals, macroalgae, and others). These observations were made in consecutive years at Caribbean, Kenyan and Great Barrier Reef sites. We assume that the state of the reef after one year follows a Dirichlet distribution with parameters dependent on the current state of the reef. These parameters are estimated using a local linear estimator with cross-validation bandwidth. These estimates are then used in a transition equation to obtain the stationary distribution of reef composition. The stationary distributions for the Caribbean and Great Barrier reef appear very different, in accordance with biological knowledge. These stationary distributions correspond to the dynamic equilibria for the two regions, if conditions remain as they are now. In addition to making predictions, our semi-parametric models provide a summary of the major features of reef dynamics, which more mechanistic models should be able to reproduce.

Keywords: Compositional data, local linear estimation.

The Evolution of Public Health in European Countries based on Stochastic Modeling Techniques

George Matalliotakis

Technical University of Crete and Technological Educational Institute of Crete

In this study we present demographic characteristics of some European Countries mainly focusing on the different mortality features of both sexes between male and female population during the last century. The life table data of several European countries are analyzed and examined. From the data analysis are derived important conclusions about the different characteristics of mortality between female and male population and indicators about the increase of average life and the variation of infant mortality. A modeling approach to Life Table Data sets is applied. The method is based on a stochastic methodology and the derived first exit time probability density function. The Health State Function $H(t)$ of the population is modeled as the mean value of the health states of the individuals and is supposed to be at a specific level at the time of birth and then changes and gradually decreases to a zero level at the time of death. The form of the density function includes the high level of deaths occurring the first years of the childhood.

Life Table Data Analysis and Mortality Prediction Applying Dynamic Modelling

George Matalliotakis^{1,3}, Christos H. Skiadas², Yiannis Dimotikalis³

¹Technical University of Crete, ²ManLab, Technical University of Crete,

³Technological Educational Institute of Crete

In the present paper with the formulation of a dynamic stochastic model is analyzed the life table data of Sweden and the appropriate charts are formed. The model expresses the human life table data by using the first-passage-time theory for a stochastic process, called Extended Quadratic Model. The model is applied to the mortality data in Sweden. The application of the model to the life table data performed by the process of non-linear regression using the following formula of the probability density function $g(t)$:

$$g(t) = k(t)^{-3/2} e^{-(l-(bt)^c)^2/2}$$

The study shows that with the dynamic modeling we can model with great success the life table data, the infant mortality and determine the age with the maximum death rate.

We present a method for forecasting mortality using the Extended Quadratic Model. Furthermore, the results derived are discussed

Author Index

Abbas, Karim	17, 104	Assi, N'Guessan	193
Abdallah, Mkhadri	193	Atalik, Gültekin	153
Abdelkader, Tatachak	205	Atilgan, Yasemin Kayhan	113
Abedieh, Jasmin	96	Auray, Stéphane	19
Acosta, Lesly M.	8	Avros, Renata	19
Afanasyeva, L.G.	9	Aziza, Ahmedou	9
Aghiles, Imine	146	Azriel, David	20
Aissani, Amar	10	Baan, Maryam	21
Aïssani, Djamil	17, 104	Badagian, Ana Laura	22
Akkartal, Erkut	154	Baek, Young Hwa	108
Aksamit, Anna	53	Baey, Charlotte	22
Al-Awadhi, F. A.	11	Baghdali-Ourbih, Latifa	23
Alawadhi, Fahima	12	Bahraoui, Zuhair	24
Alawadhi, Shafiqah	12	Balakrishnan, Narayanaswamy	1, 84
Albering, Heino	13	Baran, Sándor	24
Albert, Carlo	137	Barbu, Vlad Stefan	25
Albert, Mélisande	13	Barone-Adesi, Giovanni	26
Alexandre-Cortizo, E.	46, 86, 172	Barros, Henrique	170
Alfredo, Félix	168	Barton, Maria	145
Al-Khouri, Ritab	11	Bartosova, Jitka	26
Alonso, J.	168	Bashtova, E.E.	9
Alonso, José	167	Basna, Rani	27
Amârioarei, Alexandru	14	Bazarova, Alina	27
Ammou, Samir Ben	82	Beatriz, Lacruz	125
Anna, Soos	201	Bécue-Bertaut, Mónica	121
Anson, Jon	14	Behtari, Behnam	28
Antonov, Valery	15, 224	Belarbi, Soumia	61
Anwar, Masood	16	Belles-Sampera, Jaume	29
Anwar, Muhammad Rizwan	17	Belzunce, Félix	29
Aoumer, Sabrina	17	Benth, Fred Espen	30
Arriaga-Jordán, Carlos M.	210	Bercher, Jean-François	30
Asiimwe, Innocent	18		

Berghe, Greet Vanden	116	Capilla, Carmen	45
Berkes, Istvan	27, 178	Carrión, Paúl	156
Bersimis, Sotiris	31	Carro-Calvo, L.	46
Besnik, Pumo	9	Carty, Dillon	227
Bettaeib, Jihene	206	Caruana, Mark Anthony	185
Bilbao, M. N.	32	Casanova-Mateo, C.	32
Bina, Vladislav	26	Castañer, A.	47
Bobbio, Andrea	132	Castro, Inmaculada T.	150
Bogolubov, N.N.(Jr.)	33	Castro, Luis	34
Bohdalová, Mária	33	Ceranka, Bronislaw	47, 48
Boj, Eva	57	Chabchoub, Habib	123
Bolancé, Catalina	24	Chacón, José E.	48
Bonnéry, Daniel	34	Chambaz, Antoine	207
Borges, Ana	34	Chang, Yuan-chin Ivan	49
Bouaziz, M.-F.	35	Chen, Jie	50
Boucher Bergeron, Marie Pier	72	Chen, Yuting	50
Boukhetala, Kamal	82	Cheung, Karen Siu Lan	101
Bouret, Yann	13	Chimitova, Ekaterina	51, 52
Bouveyron, Charles	36	Chiou, Jeng-Min	52
Bozeman, James R.	37	Chlif, Sadok	206
Brabec, Marek	38	Chough, KC	107
Braumann, Carlos A.	126	Choulli, Tahir	53
Breidt, F. Jay	34	Christou, Vasiliki	54
Brekhman, Grigori I.	39	Chuang, Rwei-Ju	54
Brogi, Marina	41	Chukwu, W.I.E.	55
Bruno, John F.	229	Clancy, Damian	229
Bugrien, J.	159	Clausel, M.	55
Bulinskaya, Ekaterina	42	Coppola, Mariarosaria	56
Bystrov, Alexander	43	Coquet, François	34
Cacciolatti, Luca	139	Corberán-Vallet, Ana	187
Calcagn, Antonio	43	Corlosquet-Habart, Marine	56
Campillo, Fabien	44	Costa, Teresa	57
Campos, Viviane S.M.	172	Cournède, Paul-Henry	22, 50, 138, 208
Cañadas, Álvaro	44		

Crippa, Franca	58	Duong, Tarn	48
Cséfalvaiová, Kornélia	58	Ebeling, Marcus	72
Cuadra, L.	46	Ebrahimi, V.	75
Cuadra-Rodríguez, L.	172	Elias, Ould Said	205
Cuadras, Carles M.	148	El-Khatib, Youssef	76
D'Amato, Valeria	56, 59	Elliott, Robert J.	76
D'Amico, Guglielmo	41, 60, 61, 179	Engström, Christopher	77
Dahmani, Abdelnasser	23	Escabias, Manuel	209
Dahmani, Zoubir	61	Escobar, Luis A.	222
D'Amico, G.	177	Espinosa, José A.	210
Daniel, Garcés	156	Espinoza, José J.	210
Dávalos, José L.	210	Esteve, Anna	57
Daveloose, C.	62	Falivab, M.	175
Daveloose, Catherine	62	Farhadzadeh, E.M.	77
Davison, Matt	151	Fazekas, István	78, 79
Dedu, Silvia	64, 177	Feigin, Paul D.	20
Delibašić, Boris	217	Fernández, Lexuri	80
Demin, Victor	51	Fernández-Aguirre, Karnele	143
Deng, Jun	64	Fersi, Kmar	82
Di Biase, Guiseppe	41, 60	Feyter, Tim De	116
Diaz-Iriberry, I.	88	Fiala, Tomas	68, 83
Dimotikalis, Y	230	Filippov, V.V.	69
Djamah, Tounsia	99	Fokianos, Konstantinos	54
Djamil, Aissani	146	Foxx, Irina L.	94
Djehiche, Boualem	65	Forbelska, Marie	26
Dmitrieva, Ludmila A.	66, 67	Fortiana, Josep	57
Dondeynaz, Céline	133	Franco, Manuel	84
Doray, L.G.	97	Freire, Flavio	130
Dos Santos, Jozemar Pereira	173	Fromont, Magalie	13, 85
Dotlacilova, Petra	68, 83	Fu, James C.	134, 135
Dubinín, N.E.	69	Fullerton, Ken	145
Dubkova, Natalja	70	Fung, Eric S.	76
Dubovetska, Iryna	71	Gadag, Veeresh	86

Gahtori, Maneesha	195	Grammont, Franck	13
Galanova, Natalya	52	Gregorio, Alessandro De	63
Gallo-Marazuela, D.	32, 86, 172	Greguš, Michal	13, 33
Gamboa, Diana Carolina	87	Groll, Andreas	96
Gamiz, M.L.	87	Groparu-Cojocaru, I.	97
Gao, W.	200	Gubceac, Ghennadii	171
García, G. Maria J.	100	Guerbyenne, Hafida	97
Garcia-Santiago, C. A.	88	Guerry, Marie-Anne	116
Garg, Lalit	145	Guess, Frank M.	227
Gavrilov, L.A.	1, 90, 91	Guessoum, Zohra	98
Gavrilova, N.S.	1, 90, 91	Guillén, Montserrat	29, 167
Geffray, Segolen	91	Gunay, Suleyman	113, 173
Gershikov, Evgeny	92	Haberman, Steven	3, 59
Geystelen, J Van	211	Haddadene, H. Ait	127
Ghettab, Sara	98	Hadi, Ali S.	98
Gillermo, Saura González	168	Hamadouche, Djamel	10
Gillespie, Jennifer	145	Hammouche, Sofiane	99
Gil-Lopez, S.	88	Hamon, Thierry	77
Gil-López, S.	184	Hashorva, Enkelejd	188
Girardin, Valérie	25, 93	Hatemi-J, Abdunasser	76
Gismondi, Fulvio	94	Heo, Giseon	100
Glaz, Joseph	2, 50, 220	Hernández, G. Javier J.	100
Glidden, Eva	37	Hernández, R. José G.	100
Glukhova, Irina Yu.	94	Horiuchi, Shiro	101
Gomes, Dulce	126	Horová, Ivanka	119, 127, 216
Gonçalves, A. Manuela	165	Horvath, Lajos	27
Góngora, Sergio F.	210	Hruschka, Harald	102
Gonzaga, Marcos	130	Hudecová, Šárka	103
González, Arturo	210	Hurd, Tom	104
González, Nelfi	222	Husson, François	121
Gonzalez-Gonzalez, A.	88	Ibrahim, Sidi Zakari	193
Gozgor, Giray	95	Ildiko, Somogyi	201
Graczyk, Malgorzata	47, 48	Inui, Koji	112

Iranipour, S.	75	Khedher, Asma	30
Issaadi, Badredine	104	Kim, Ho-Seok	108
Ivanova, Alla	105	Kitaeva, Anna V.	115
Ivin, Evgeny A.	106	Klutchnikoff, Nicolas	19, 91
Ivnitskiy, V.A.	106	Kocsis, Orsolya	115
Jacques, Julien	36	Koláček, Jan	119, 127, 216
Janssen, Jacques	56, 94	Kollo, Tõnu	116
Jeanblanc, Monique	53	Komarudin	116
Jean-Marie, Marion	9	Komornik, Jozef	118
Jiang, Bei	107	Komornikova, Magda	118
Jiménez-Fernández, S.	46	Konar, Ondrej	38
Jin, Hee-Jeong	108	Konecna, Katerina	119
Jizba, Petr	107	Konovalov, Mikhail G.	119
Joannides, Marc	44	Korbel, Jan	107
Jolalpa, José L.	210	Koshkin, Gennady M.	94
Joly, Pierre	108	Kosolapov, Samuel	92, 120
Joo, Yoohyung	109	Kostov, Belchin	121
Jørgensen, Bent	110	Kovalenko, Anatoly	15
Jovanović, Miloš	217	KRIT, Makram	122
Kadankova, Tetyana	110	Krzeminski, Mark	222
Kaděrová, Jana	217	Ktari, Raida	123
Kaiser, Regina	22	Kundu, Debasis	84
Kaishauri, Tinatin	111	Kuperin, Yuri A.	66, 67
Kakubava, Revaz	111	Kurbatskiy, Alexey N.	106
Kalamatianou, Aglaia	111	Kutoyants, Yury A.	4
Kanunikov, Igor E.	66	Kuttykrishnan, A.P.	124
Karácsony, Zsolt	78	Labiod, Lazhar	125
Karimzadeh, R.	75	Lacruz, Beatriz	98
KARIYA, Takeaki	112	Lagarto, Sandra	126
Kasanicky, Ivan	38	Lahoz, David	125
Kernane, Tewfik	113	Lajdová, Dagmar	127
Kessira, Abderrahim	97	Lamamri, A.	127
Kharin, Alexey	114	Landa-Torres, I.	88, 184
Khedher, A.	62	Langhamrová, Jitka	58, 68, 83

Larik, Noor.M.	128	Manfrin, Andrea	139
LAURENT, Béatrice	85	Manita, Anatoly	140
Lee, Hee Yeon	109	Manita, Larisa	140
Lee, Si-woo	108	Mansouri, Rachid	99
Lefèvre, Claude	128	Marinescu, Ionela	141
Lenca, Philippe	56	Marinela	
Leoncini, Emanuele	129	Mármol, M.	47
Leoni-Aubin, Samuela	206	Márquez, M. Dolores	142
Lerasle, Matthieu	85	Marshall, Adele H.	111, 142
Levendovszky, János	196	Martín-Arroyuelos, Ana	143
Lhote, Loïck	93	Martínez, Francisco E.	210
Li, Yin	100	Martínez, I. Pedraza	168
Lifshits, Mikhail	188	Martínez-Riquelme, Carolina	29
Lillo, Rosa E.	181	Mase, Shigeru	144
Lima, Everton	130	Matalliotakis, G.	230
Limon, F. Duran	88	Mateu, Jorge	147
Lindqvist, B.H.	87	Mazouch, Petr	145
Liu, Shiang-Tai	131	Mazzoleni, Marcella	58
Lombardi, Luigi	43, 132	McClanahan, Tim	229
López, Alberto	132	McClean, Sally	145
Lou, Wendy	135	Medici, Giovanna De	63
Luna, América A.	210	Mejía, Raúl	156
Luz, Maksym	136	Mekkaoui, A.	147
Mateo Pedro, M.	125	Mellah, Rabah	99
Claramunt, M.M.	47	Melo, Carlos E.	147
Macci, Claudio	137	Melo, Oscar O.	147, 148
Machac, David	137	Melo, Sandra E.	148
Malefaki, Sonia	138, 208	Memmedli,	149
Malin, Prakash	139	Memmedaga	
Malouche, Dhafer	207	Mendes, Mehmet	154
Maly, Marek	38	Menezes, Raquel	181
Manca, Raimondo	41, 56, 60, 63, 94, 194	Mercier, Sophie	150
Mancini, Loriano	26	Michalopoulou, K.	202
Mandelbaum, Avishai	20	Mielkie, Melissa	151
		Milne, Erin	37

Minin, Alexey S.	67	Novák, Petr	162
Minkova, Leda	152	Novitsky, N.N.	162
Mizrak, Gökem	153	Nucci, Massimo	132
Moctezuma, Georgel	210	Núñez-Antón, Vicente	166
Modroño-Herrán, Juan Ignacio	143	Nyende, Joseph	164
Moklyachuk, Mykhailo	71, 136	Nyul, Balázs	165
Molina, Manuel	153	Ohta, Yasutaka	221
Molinero, Cecilio Mar	139, 141, 174	Oliveira, Raquel	165
Mollaogullari, Ahmet	154	Orbe, Jesus	166
Montenegro, Álvaro Mauricio	87	Ornelas, Arelly	167
Monter, Rosario	154	Ortega, EM.	168
Moodjom, Roland Mark	155	Ortega, Eva Maria	167
Moraes, Alvaro	156	Ortega-Quiles, E.	168
Moran-Zuloaga, Daniel	156	Osorio, Felipe	110
Moreno, César Carmona	133	Ostrovskaya, Darya	169
Mota, Manuel	153	Ouellette, Nadine	101
Mukhopadhyay, Barun Kumar	157	Ourbih-Tari, Megdouda	23
Mulero, Julio	29	Ozdemir, Ozer	149
Muñoz, M. Pilar	8, 32, 142	Pacchiarotti, Barbara	137
Muradaliev, A.Z.	78	Paes, Neir A.	170
Murugesan, P.	158	Paes, Neir Antunes	173
Mwitondi, K.	159	Paladi, Florentin	171
Nadif, Mohamed	125	Papaioannou, Takis	31
Nagih, A.	127	Pastore, Massimiliano	132
Nasiri, Mahdi	21	Pastor-Sánchez, A.	86, 172
Nassar, Hiba	65	Pelikan, Emil	38
Negri, Ilia	159	Peña, Daniel	22
Ng, H.K.T.	200	Pereira, André G.C.	172
Nguyen, Thai H.	160	Pérez-Marín, Ana M.	24
Nguyen3, Cuong	118	Pérez-Palomares, Ana	98
Nicholson, Timothy	37	Petroni, F.	177
Niemiro, Wojciech	160	Petroni, Filippo	61, 63
Nishihara, Michi	161	Pettere, Gaida	116
		Picard, Philippe	128

Piscopo, Gabriella	59	Rinott, Yosef	180
Pistone, Giovanni	212	Robine, Jean-Marie	5, 101
Poirot, E P	211	Rocha, Lisandra	181
Polat, Esra	173	Rodríguez, Joanna	181
Portilla-Figueras, A.	46, 86	Romero, Paola	156
Portilla-Figueras, J. A.	184	Roschin, D.O.	182
Portillo, Fabiola	141, 174	Rosin, Klemens	182, 187
Porvázsnnyik, Bettina	79	Rouvière, Laurent	19
Potia, V.	175	Ruffa, Suela	212
Prattico, F.	177	Ruiz, José M.	29
Preda, Cristian	14	Russolillo, Maria	59
Preda, Vasile	177	Saadati, Elahe	21
Prieto, L.	86	Sabgayda, T.	189
Puga, Jorge López	133	Sabgayda, Tamara	105
Puig, Pere	4	Sabgayda, TP	182
Quang, Vu van	15, 224	Sachlas, Athanasios	31
Queiroz, Bernardo	130	Sadki, Ourida	98
Rabah, Medjoudj	146	Saenko, Viacheslav	183
Rade, Diana	44	Safi, Samir K.	184
Rafik, Medjoudj	146	Salah, Afif Ben	206
Ramesh, B M	139	Salah, Afif Ben	207
Ramírez-Cobo, Pepa	181	Salcedo-Sanz, S.	32, 46, 86, 88, 172, 184
Ramos, Alfonso	153	Salha, Raid B.	185
Raseta, Marko	178	Salvi, Giovanni	41
Rasulova, M.Yu.	33	Sánchez, Josep A.	132
Regnault, Philippe	178, 179	Sanni, S.S	55
Reichert, Peter	137	Sant, Lino	185
Rejchel., Wojciech	160	Santalova, Diana	186
Reynaud-Bouret, Patricia	13, 85	Santolino, Miguel	29
Řezáč, Martin	180	Santonja, Francisco J.	187
Riaz, Muhammad	227	Scarsini, Marco	6
Ricardo, Osés	168	Schekoldin, Vladislav	51
Rodríguez		Scherer, Matthias	80
Riegelning, Judith	182, 187	Schmuland, Byron	100

Schwierz, Cornelia	182, 187	Soriano, Pilar	142
Seco, Luis	188	Sousa, Inês	34, 181
Sedigheh Mirzaei, S.	152	Soós, Anna	201
Seleznev, Oleg	188	Spencer, Matthew	229
Semushin, Innokentiy V.	189	Spingola, Andrea	202
Semyonova, V.	189	Stamatopoulou, G.	202, 205
Semyonova, VG	182	Stehlík, Milan	24
Semyonova, Viktoria	105	Stepanova, Natalya V.	115
Sene, Massamba	190	Stoimenova, Vessela	203
Sengupta, Debasis	152	Suárez-Llorens, Alfonso	29
Şentürk, Sevil	153	Suknović, Milija	218
Ser, J. Del	32, 88, 184	Sviben, Marijan	203
Serban, Florentin	64	Svishchuk, Mariya	204
Shakhgildyan, Ksenia	192	Swishchuk, Anatoliy	204
Shaptiley, Maria A.	66, 67	Symeonaki, M.	202, 205
Shefrin, Hersh	26	Taleb, Samira	10
Shonia, Otar	111	Teixeira, Cristina	170
Sikolya, Kinga	24	Tempone, Raul	156
Silvestrov, Dmitrii	6, 194	Tesárková, Klára	145
Silvestrov, Sergei	77	Hulíková	
Silvestrova, Evelina	194	Tishaboev, I.A.	33
Šimková, Martina	58	Toledano-Kitai, Dvora	19
Simpach, Ondrej	194	Toma, Aida	206
Singh, S. B.	195	Toumi, Amine	206, 207
Sipos, I. Róbert	196	Trapani, Lorenzo	59
Siu, Tak Kuen	76	Trevezas, Samis	22, 50, 138, 208
Skiadas, Charilaos	197, 199	Tsyganova, Julia V.	189
Skiadas, Christos H	197, 199, 230	Turkan, Semra	173
Skovikov, Anatoli G.	189	Valderrama, Mariano J.	209
Slovesnov, Alexandr V.	106	Valdivia, Arturo	210
Smetanin, Nikolai M.	66	Valdivia, Arturo G.	210
So, H.Y.	200	Valge, Marju	116
Soffer, Avi	19	Vallejos, Ronny	110
		Valverde, Irène Larramendy	44

Vanmaele, M.	62	Y.Z., Farzaliev	78
Vanteyeva, O.V.	162	Yacine, Ferrani	205
Vas, Renáta	78	Yamamoto, Kyohei	221
Vasconcelos, Rosa M.	165	Yamamua, Yoshiro	112
Vatolin, N.A.	69	Yáñez, Sergio	222
Vedel, B.	55	Yarkulov, Boli	223
Vélez, Alejandra	210	Yarovaya, Elena	223
Ventura, Giuseppina	63, 94	Yi, Qilong	134
Vicario, Grazia	212	Yi, Qilong	135
Vilanova, Pedro	156	Yildiz, Munevvere	149
Vilčko, František	44	Yoshimoto, Kouji	144
Vimond, Myriam	91	Young, Timothy M.	227
Víšek, Jan Ámos	214	Yurij, Saenko	183
Vivo, Juana-Maria	84, 216	Yuryev, A.A.	69
Volkovich, Zeev	19	Zagaynov, Artem	15, 224
Vopatová, Kamila	216	Zaiats, Vladimir	225
Vořechovský, Miroslav	217	Zaigrajew, Aleksander	225
Vukičević, Milan	217	Zalewska, Marta	160
Vvedenskaia, Elena	218	Zamaï, Eric	35
Wan, Alan	219	Zaman, Babar	226, 227
Wang, Binbin	220	Zambrano, Carlos	44
Wang, K.	159	Zemlyanova, Elena	105
Wang, Xiao	220	Zeng, Yan	227
WANG, Zhu	112	Zenga, Mariangela	58, 111, 142
Ware, Antony	220	Zhang, Xinyu	219
Watanabe, Erika	221	Zinchenko, Nadiia	228
Watanabe, Norio	221	Zoiab, M. G.	175
Xu, Xiaojian	222	Zychaluk, Kamila	229
Xu, Y.	147		