Book of Abstracts

17th Applied Stochastic Models and Data Analysis
International Conference with
Demographics Workshop

ASMDA2017

Editor

Christos H Skiadas

6 - 9 June 2017

London, UK

Imprint
Book of Abstracts of the 17th Applied Stochastic Models and Data Analysis International Conference with the 6th Demographics Workshop
London, UK: 6-9 June, 2017
Published by: ISAST: International Society for the Advancement of Science and Technology.
Editor: Christos H Skiadas

Book ISBN: 978-618-5180-22-5

Preface

It is our pleasure to welcome the guests, participants and contributors to the International Conference (ASMDA 2017) on Applied Stochastic Models and Data Analysis and (DEMOGRAPHICS2017) Demographic Analysis and Research Workshop.

The main goal of the conference is to promote new methods and techniques for analyzing data, in fields like stochastic modeling, optimization techniques, statistical methods and inference, data mining and knowledge systems, computing-aided decision supports, neural networks, chaotic data analysis, demography and life table data analysis.

ASMDA Conference and DEMOGRAPHICS Workshop aim at bringing together people from both stochastic, data analysis and demography areas. Special attention is given to applications or to new theoretical results having potential of solving real life problems.

ASMDA 2017 and DEMOGRAPHICS 2017 focus in expanding the development of the theories, the methods and the empirical data and computer techniques, and the best theoretical achievements of the Applied Stochastic Models and Data Analysis field, bringing together various working groups for exchanging views and reporting research findings.

We thank all the contributors to the success of these events and especially the authors of this Proceedings Book. Many thanks to the honorary guest Gilbert Saporta and the Colleagues contributed in his special session on data analysis. Special thanks to the Plenary, Keynote and Invited Speakers, the Session Organisers, the Scientific Committee, the ISAST Committee, Yiannis Dimotikalis, Aristeidis Meletiou, the Conference Secretary Mary Karadima, and all the members of the Secretariat.

May 2017

Christos H. Skiadas
Conference Chair

ASMDA Conferences and Organizers

1st ASMDA 1981 Brussels, Belgium. Jacques Janssen

2nd ASMDA 1983 Brussels, Belgium. Jacques Janssen

3rd ASMDA 1985 Brussels, Belgium. Jacques Janssen

4th ASMDA 1988 Nancy, France. J. Janssen and Jean-Marie Proth

5th ASMDA 1991 Granada, Spain. Mariano J. Valderrama

6th ASMDA 1993 Chania, Crete, Greece. Christos H Skiadas

7th ASMDA 1995 Dublin, Ireland. Sally McClean

8th ASMDA 1997 Anacapry, Italy. Carlo Lauro

9th ASMDA 1999 Lisbon, Portugal. Helena Bacelar-Nicolau

10th ASMDA 2001 Compiègne, France. Nikolaos Limnios

11th ASMDA 2005 Brest, France. Philippe Lenca

12th ASMDA 2007 Chania, Crete, Greece. Christos H Skiadas

13th ASMDA 2009 Vilnious,Lithouania. Leonidas Sakalauskas

14th ASMDA 2011 Rome, Italy. Raimondo Manca

15th ASMDA 2013 Mataró (Barcelona), Spain. Vladimir Zaiats

16th ASMDA 2015 Piraeus, Greece. Sotiris Bersimis

17th ASMDA 2017 London, UK. Christos H Skiadas

SCIENTIFIC COMMITTEE

Plenary/Keynote Talks For ASMDA Conference

In celebration of Gilbert Saporta's 70th birthday and in honour of his contributions to Applied Statistics and Data Analysis and his support to ASMDA activities

Gilbert Saporta
Emeritus Professor of Applied Statistics
Conservatoire National des Arts et Métiers (CNAM)
Paris, France

N. Balakrishnan
Department of Mathematics and Statistics
McMaster University
Hamilton, Ontario, Canada

Robert J. Elliott
Haskayne School of Business,
University of Calgary, Canada and
Centre for Applied Financial Studies,
University of South Australia,
Adelaide, Australia

Sally McClean
School of Computing and Information Engineering
Ulster University
Coleraine
Northern Ireland

Fabrizio Ruggeri
CNR IMATI
Via Bassini 15
Milano, Italy

Anatoliy Swishchuk
Department of Mathematics and Statistics
University of Calgary, Canada

P.-C.G. VASSILIOU
Department of Statistical Sciences,
University College London, UK

For Demographics Workshop

Jean-Marie Robine
Université Montpellier 2, Place Eugène Bataillon
Montpellier, France

Rebecca Kippen
Rural Health,
Monash University
Victoria, Australia

Contents **Page**

BOOK OF ABSTRACTS

Applied Stochastic Models and Data Analysis
ASMDA 2017 & DEMOGRAPHICS 2017

Plenary and Keynote Talks

Malliavin Calculus in a Binomial Framework

Samuel N. Cohen[1], Robert J. Elliott[2,3], Tak Kuen Siu[4]
[1]Mathematical Institute, University of Oxford, United Kingdom,
[2]Haskayne School of Business, University of Calgary, Canada,
[3]Centre for Applied Financial Studies, University of South Australia, Australia,
[4]Dept of Actuarial Studies and Centre of Financial Risk, Faculty of Business and Economics, MacQuarie University, Australia

The binomial model is a standard framework used to introduce risk neutral pricing of financial assets. Martingale representation, back- ward stochastic differential equations and the Malliavin calculus are difficult concepts in a continuous time setting. This paper presents these ideas in the simple, discrete time binomial model.

Projecting Life Expectancy: A Global History

Rebecca Kippen
School of Rural Health, Monash University, Australia

This paper covers the history of life expectancy projections. It starts with nineteenth-century debates on whether and why the ‘duration of human life’ was increasing, then details twentieth-century projections and methods that repeatedly underestimated subsequent declines in mortality. Finally, I discuss recent methodological advances in projections and speculations about the future course of life expectancy.

Keywords: population projections, life expectancy, demographic methods, history.

Markov and semi-Markov Models for Health and Social Care Planning

Sally McClean
Computer Science Research Institute, Ulster University, United Kingdom

In the 1980s, a geriatrician at St George's Hospital in London, Professor Peter Millard, built a novel database of patient length-of-stay (LOS) and showed that patient LOS could be modelled by simple Markov models, reflecting essential features of patient behaviour. Such Coxian phase-type models (essentially k-state progressive Markov models) have subsequently been shown to work well for a range of settings and scales, including hospitals, community care, emergency services and patient activity recognition. These models can be extended to predict individual behaviour, to assess resource needs and costs, and are intuitively appealing as they conceptualise patient progression, for example, through acute care, treatment, and rehabilitation. More accurate predictions require the inclusion of covariates to represent differences between patients, through for example, modelling transition rates in terms of covariates, conditional phase-type distributions, or phase-type survival trees. Such covariates may (i) be available at admission (e.g. age, admission method, diagnosis); (ii) be ongoing (e.g. treatment, sensor readings) or (iii) be external (e.g. resource constraints).

Using Markov and semi-Markov models, based on these ideas, we have developed an integrated modelling framework for patient care, which identifies patient pathways, based on covariates such as age, gender, or diagnosis, and multiple outcomes, such as discharge to normal residence, nursing home, or death. We thus model systems that encompass the whole care process and include multiple pathways, containing various states and phases, some of which are absorbing states; we then use this integrated model to facilitate planning of services. In addition we assume Poisson, or more complex, distributions of admissions to the system. A model of the whole care system is thus developed and results obtained for moments of length of stay and numbers of patients in different states of the system at any time. We here focus on probabilities, moments of patient numbers in different parts of the system, length of stay and cost expressions for such Markov and semi-Markov models with applications to health and social care services.

Recent Advances in Adversarial Risk Analysis

Fabrizio Ruggeri
CNR IMATI, Italy

Adversarial Risk Analysis (ARA) is an emergent paradigm for supporting a decision maker who faces adversaries in problems in which the consequences are random and depend on the actions of all participating agents. ARA aims at providing one-sided prescriptive support to one of the intervening agents, the Defender (D, she), based on a subjective expected utility model treating the adversary's decisions as uncertainties. To do so, we model the adversary's (A, Attacker, he) decision making problem and, assuming that he is an expected utility maximiser, try to assess his probabilities and utilities. We can consequently forecast his optimal action. In the talk we will first provide a general overview on ARA and then we will present two recent works. In the first we will discuss some adversarial issues in reliability about acceptance sampling and exponential life testing whereas in the second we introduce the notion of Adversarial Hypothesis Testing which enlarges standard hypothesis testing by including an adversary which aims at distorting the relevant data processes to confound a decision maker. We provide an ARA approach to this problem and illustrate its usage in spam detection.

50 Years of Data Analysis: from EDA to Predictive Modelling and Machine Learning

Gilbert Saporta
CEDRIC-CNAM, France

In 1962, J.W. Tukey wrote his famous paper “The future of data analysis” and promoted Exploratory Data Analysis (EDA), a set of simple techniques conceived to let the data speak, without prespecified generative models. In the same spirit J.P. Benzécri and many others developed multivariate descriptive analysis tools. Since that time, many generalizations occurred, but the basic methods (SVD, k-means, …) are still incredibly efficient in the Big Data era.
On the other hand, algorithmic modelling or machine learning are successful in predictive modelling, the goal being accuracy and not interpretability. Supervised learning proves in many applications that it is not necessary to understand, when one needs only predictions.
However, considering some failures and flaws, we advocate that a better understanding may improve prediction. Causal inference for Big Data is probably the challenge of the coming years.

Keywords: exploratory data analysis, machine learning, prediction.

Financial Mathematics: Historical Perspectives and Recent Developments

Anatoliy Swishchuk
Dept of Mathematics and Statistics, University of Calgary, Canada

This talk is devoted to the diverse historical perspectives of financial mathematics and its recent developments including energy finance, systemic risk and algorithmic and high-frequency trading theories.

Laws of Large Numbers for Non-Homogeneous Markov Systems

P.-C. G. Vassiliou
Dept of Statistical Sciences, University College London, United Kingdom

One of the most celebrated theorems in Probability theory are the Laws of Large numbers. In the present we study Laws of Large numbers for non-homogeneous Markov systems (NHMS). We provide theorems for convergence in the mode of L^2 and also for the almost surely mode. Cyclic non-homogeneous Markov systems is a category NHMSs with its own interst due to the important applications that exist. In the present we also study Laws of Large numbers for non-homogeneous Markov systems under cyclic behavior (Cyc-NHMS). We provide theorems for convergence in the mode of L^2 and also for the almost surely mode. Illustrated applications are also provided for the above described theorems.

Invited Talks

Risk measures based on Option Prices and Changes in the Jump Process of Asset Returns

Giovanni Barone Adesi
Università della Svizzera italiana, Switzerland

The derivative of the option price with respect to the strike price identifies the probability of the tail area whenever an Arrow-Debreu representation of prices is possible. This general result identifies Value at Risk and, jointly with the option price, expected shortfall. It does not require inferences on the extreme tail of the distribution. For short time horizons, the effect of the change from pricing to physical measure becomes negligible.
The ratio of expected shortfall over Value at Risk changes through time, pointing to changes in the form of the distribution of returns perceived by investors under the pricing measure. Empirical tests investigate whether that is due to changes in the physical distribution or risk preferences.

The Likelihood Ratio Test for the Equality of Mean Vectors when the Covariance Matrices are Block Compound Symmetric

Carlos A. Coelho
Centro de Matemática e Aplicações (CMA-FCT/UNL) and Mathematics Dept, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal

The likelihood ratio test for the equality of mean vectors when the covariance matrices are assumed just positive-definite is a common test in Multivariate Analysis, but similar likelihood ratio tests are not available in the literature when the covariance matrices are assumed to have some common given structure, or, if available, they usually entail very long and tedious derivations. Since the block compound symmetric covariance structure may be an adequate covariance structure in a large number of situations and in a large number of models, the development of a likelihood ratio test for the equality of mean vectors when the covariance matrices are assumed to have a block compound symmetric structure is of great interest.
The author presents a straightforward derivation of the likelihood ratio statistic for this test and shows how from the approach followed it is possible to obtain, for some cases, a finite simple representation for both

the probability density and the cumulative distribution functions of the statistic. For the other cases, given the intractability of the expressions for these functions, very sharp near-exact distributions are developed. These near-exact distributions allow for an easy computation of quantiles and p-values and are shown to lie very close to the exact distribution, even for very small sample sizes. Furthermore, they are shown to be asymptotic not only for increasing sample sizes but also for increasing number of variables and populations involved.

Keywords: Characteristic function, Likelihood ratio statistic, near-exact distributions, Product of Beta random variables.

Acknowledgements: *Research partially supported by FCT–Fundação para a Ciência e Tecnologia (Portugal), through project UID/MAT/00297/2013 – Centro de Matemática e Aplicações (CMA-FCT/UNL)*

Penultimate Approximations in Extreme Value Theory and Reliability of Large Coherent Systems

M. Ivette Gomes[1], Paula Reis[2], Luísa Canto e Castro[3], Sandra Dias[4]
[1]CEAUL and DEIO, Faculdade de Ciências, Universidade de Lisboa, Portugal, [2]CEAUL and MAT, EST-IPS, Portugal, [3]CEAUL, Universidade de Lisboa, Portugal, [4]Pólo CMAT-UTAD and CEMAT, Universidade de Trás-os-Montes e Alto Douro, Portugal

The rate of convergence of the sequence of linearly normalized máxima /minima to the corresponding non-degenerate *extreme value* (EV) limiting distribution for máxima /minima is a relevant problem in the field of *extreme value theory*. In 1928, Fisher and Tippett observed that, for normal underlying parents, if we approximate the distribution of the suitably linearly normalized sequence of maxima not by the so-called Gumbel limiting distribution, associated with an *extreme value index* (EVI) $\xi = 0$, but by an adequate sequence of other EV distributions with an EVI $\xi_n = o(1) < 0$, the approximation can be asymptotically improved. Such approximations are often called *penultimate approximations* and have been theoretically studied from different perspectives. Recently, this same topic has been revisited in the field of *reliability*, where any coherent system can be represented as either a *series-parallel* or a *parallel-series* system. Its lifetime can thus be written as the minimum of maxima or the maximum of minima. For large-scale coherent systems, the possible non-degenerate EV laws are thus eligible candidates for the finding of adequate lower and upper bounds for such system's reliability. However, just as mentioned above, such non-degenerate limit laws are better approximated by an adequate penultimate distribution in most situations. It is thus sensible to assess both theoretically and through Monte-Carlo simulations the gain in accuracy when a penultimate

approximation is used instead of the ultimate one. Moreover, researchers have essentially considered penultimate approximations in the class of EV distributions, but we can easily consider a much broader scope for that type of approximations, and such a type of models surely deserves a deeper consideration under statistical backgrounds. A few details on these topics will be presented.

Keywords: Extreme value theory, Monte-Carlo simulation, penultimate and ultimate approximations, system reliability.

Mean First Passage Times in Markov Chains – How Best to Compute?

Jeffrey Hunter
School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, New Zealand

The presentation gives a survey of a variety of computational procedures for finding the mean first passage times in Markov chains. The presenter has recently published a new accurate computational technique (Special Matrices, 2016) similar to that developed by Kohlas (Zeit. fur Oper. Res., 1986) based on an extension of the Grassmann, Taksar, Heyman (GTH) algorithm (Oper. Res., 1985) for finding stationary distributions of Markov chains. In addition, the presenter has recently developed a variety of new perturbation techniques for finding key properties of Markov chains including finding the mean first passage times (Linear Algebra and its Applications, 2016). These procedures are compared with other well known procedures including the standard matrix inversion technique (Kemeny and Snell, 1960), some simple generalized matrix inverse techniques developed by the presenter (Asia Pacific J. Oper. Res., 2007), and the FUND technique of Heyman (SIAM J Matrix Anal. and Appl., 1995) for finding the fundamental matrix of a Markov chain. The accurate procedure of the presenter is favoured following MatLab comparisons using some test problems that have been used in the literature for comparing computational techniques for stationary distributions. One distinct advantage is that the stationary distribution does not have to be found in advance but is extracted from the computations.

Keywords: Markov chains, stationary distributions, generalized matrix inverses, perturbations, fundamental matrix, mean first passage times.

Epidemic Risk and Insurance Coverage

Claude Lefèvre
ISFA, Lyon and ULB, Belgium

This paper aims to apply simple actuarial methods to build an insurance plan protecting against an epidemic risk in a population. The studied model is an extended SIR epidemic in which the removal and infection rates may depend on the number of registered removals. The costs due to the epidemic are measured through the expected epidemic size and infectivity time. The premiums received during the epidemic outbreak are measured through the expected susceptibility time. Using martingale arguments, a method by recursion is developed to calculate the cost components and the corresponding premium levels in this extended epidemic model. Some numerical examples illustrate the effect of removals and the premium calculation in an insurance plan.

This is a joint work with P. Picard (ISFA) and M. Simon (ULB).

Event and Its Location Detection in a Wireless Sensor Network

Tapan K. Nayak
Dept of Statistics, George Washington University, USA

Wireless sensor networks, which can often be installed quickly and fairly economically, are useful for detecting threats (or events) in a region of interest. As the data received from sensor nodes contain measurement and transmission errors, interpreting the data requires appropriate statistical methods and algorithms. In particular, deciding if an event is present in the network region or not and inferring the location of the event when it is deemed present are two important decision problems. We give a statistical framework for addressing these two problems and frame them as one estimation problem. We present a solution based on the maximum likelihood method and evaluate its performance by simulation. We also describe a Bayesian approach that can be used when relevant prior distribution and loss function are available.

Keywords: Estimation error, Maximum likelihood, Parameter estimation, Sensor node, Transmission error.

Talks for Invited and Contributed Sessions

PageRank and Perron-Frobenius Theory in Analysis of Non-Negative Matrices

Benard Abola, Sergei Silvestrov
Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Sweden

Non-negative matrices are useful in many areas since their entries present data and physical meaning of real world phenomena for technology. Images, web pages, medical and financial data are some of examples. Moreover, understanding such data is challenging.
A mathematical framework to analyse non-negative matrices when performing perturbation of vertices or edges of complex evolving networks (graphs) will be presented. Properties of spectrum of the matrices derived from such graphs will be investigated. Perron-Frobenius theory will be utilised for computation of eigenvalues, eigenvectors and investigating the behavior of sub-matrices of the graphs. Numerical experiments of the perturbation of graphs structures will be demonstrated with PageRank and Power series algorithm.

Keywords: Non-negative matrices, PageRank, Power series.

Evaluation of Stopping Criteria for Ranks in Solving Linear Systems

Benard Abola, Pitos Biganda, Christopher Engström, Sergei Silvestrov
Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Sweden

Linear systems of algebraic equations arising from mathematical formulation of natural phenomena or technological processes are common. Many of these systems of equations are large, the matrices derived are mainly sparse and need to be solved iteratively. Moreover, interpretation is crucial in making decision. Bioinformatics, internet search engines (webpages) and social networks are some of the examples with large and high sparsity matrices. For some of these systems only the actual ranks of the solution vector is interesting rather than the vector itself. In this case, it is desirable that the stopping criterion reflects the error in ranks rather than the residual vector which might have a lower convergence. In this paper, we will evaluate

stopping criteria on Jacobi, successive over relaxation and power series iterative schemes. We will focus on the following criteria: residual vector, solution vector, rank correlation and rank for specific top terms. Numerical experiments will be performed and result presented.

Keywords: stopping criteria, networks, rank.

On the Behavior of the Conditional Quantile Estimator for Truncated-Associated Data

Latifa Adjoudj[1], Abdelkader Tatachak[2]
[1]Laboratory MSTD, Dept of Probability and Statistics, USTHB, Algeria,
[2]Laboratory MSTD, Faculty of Mathematics, USTHB, Algeria

The dependent data scenario is an important one in a number of applications with survival data. Due to random left truncation effect, no information is available when a subject is truncated and then, because we are only aware of subjects that we observe, the inference for truncated data is restricted to conditional estimation. In the present work, we consider the estimators of the conditional distribution and conditional quantile functions for randomly left truncated data satisfying association dependency condition. As results, we derive strong uniform consistency rates and assess the asymptotic normality of the estimators. Then the accuracy of the estimators is checked by a simulation study.

Keywords: Associated data, Kernel estimator, Quantile function, Rate of convergence, Random left truncation, Strong uniform consistency.

Asymptotic Analysis of Queueing Models by a Synchronization Method

Larisa Afanaseva
Dept of Probability Theory, Faculty of Mathematics and Mechanics,
Lomonosov Moscow State University, Russia

We consider a multi-server queueing system with a regenerative input flow X(t). Let Y(t) be the number of customers that can be served during the time interval [0, t) under assumption that there are always customers for service. Supposing Y(t) to be a strongly regenerative process we define a sequence of common regeneration points for the both processes X(t) and Y(t). The intensities of these processes can be expressed in terms of the means of their increments during the common regeneration period. Hence, the traffic rate for the system can be also obtained in these terms. If the sequence of common regeneration points can be defined in such a way that increments of Y(t) on the regeneration

period stochastically dominate the real number of customers served on this period then a theorem about conditions of the instability of the system is proved. To obtain the stability condition we additionally assume that there are two possibilities for the process Q(t) which is the number of customers in the system at time t. Namely, Q(t) tends in probability to infinity or Q(t) is stochastically bounded process as t tends to infinity. We show that the first possibility cannot take place if the traffic rate is less than one. Therefore the process Q(t) is stochastically bounded in this case. We also give some examples: multi-channel queueing system with heterogeneous servers and interruptions of the service, queueing models with priority and others.

Keywords: Regenerative flow, Synchronization, Queue, Stability.

Acknowledgment: *The research was partially supported by RFBR grant 17-01-00468.*

Stability Analysis of a Queueing Cluster Model with a Regenerative Input Flow

Larisa Afanaseva, Elena Bashtova
Dept of Probability, Faculty of Mathematics and Mechanics, Lomonosov Moscow State University, Russia

We study the stability conditions of the multi-server system in which each customer requires a random number of servers simultaneously, a random service time is identical at all occupied servers. As in the paper [1] we call this system a cluster model since it may be employed in description of modern multicore high performance clusters. Stability criterion of an M|M|s cluster model has been proved by the authors earlier and for a MAP|M|r in the recent work by E. Morozov and A. Rumyantsev. The main contribution of this work is an extension of the stability criterion to the cluster model with a regenerative input flow. The class of these flows contains the most of fundamental flows that are exploited in the queueing theory. Thus, semi-Markov, Markov-modulated, Markov arrival processes and others belong to this class. So we consider the system Reg|M|r. Our analysis is based on synchronization of the input flow X(t) and auxiliary service process Y(t) that is the number of served customers under assumption that there are always customers for service. Since service time has an exponential distribution this process turns out to be Markov-modulated one. It is shown that the intensities of X(t) and Y(t) are equal to the means of the increments of these processes on common regeneration periods for X(t) and Y(t). Moreover, Y(t) dominates the real service process (the real number of served customers). Basing on these estimates we obtain the

stability criterion, that is the same as established in [1] for a queueing system MAP|M|r.

Keywords: Regenerative flow, Cluster model, Queue, Stability.

Reference:

[1] E. Morozov, A. Rumyantsev. Stability Analysis of a MAP|M|s Cluster Model by Matrix-Analytic Method. In: Computer Performance Engineering, 63-76, Springer 2016.

Acknowledgment: *The research was partially supported by RFBR grant 17-01-00468.*

Stability Analysis of Multiserver Queueing System with a Regenerative Interruption Process

Larisa G. Afanasyeva[1], Andrey Tkachenko[2]

[1]Lomonosov Moscow State University, Dept of Mathematics and Mechanics, Russia, [2]National Research University Higher School of Economics, Dept of Applied Economics, Russia

Queueing systems in which servers may be temporary unavailable for operation arise naturally as models of many computer, communication and manufacturing systems. Service interruptions may result from resource sharing, server breakdowns, priority assignment, vacations, some external events, and others. For instance, for queueing systems with preemptive priority discipline service interruptions for the low priority customers occur when a high priority customer arrives during a low priority customer's service time. Therefore, there is significant interest in the investigation of queueing systems with service interruptions.

We consider a system with m heterogeneous servers and a common queue. The input flow $X(t)$ is assumed to be regenerative with intensity $\lambda = \lim_{t\to\infty} X(t)/t$. The preemptive repeat different service discipline is investigated, i.e. service is repeated from the beginning with different independent service time after interruption. By $B_i(t)$, $i = \overline{1,m}$ denote a distribution function of service times by the ith server and $H_i(t) = \sum_{n=1}^{\infty} B_i^{n*}(t)$. Suppose that blocking periods of the ith server $\{u_{in}^{(1)}\}_{n=1}^{\infty}$ are random variables with mean $a_i^{(1)}$. Working periods of the ith server $\{u_{in}^{(2)}\}_{n=1}^{\infty}$ are random variables with mean $a_i^{(2)}$. Sequence of availability cycles $\{u_{in}^{(1)}, u_{in}^{(2)}\}_{n=1}^{\infty}$ may consist of dependent elements, but it should have regeneration in some sense. Let $\tau_j^{(2)}$ be regeneration periods for this sequence and ν_{ij} be the number of availability cycles for the ith server during regeneration period. By $\mathrm{Q}(t)$ denote the number of customers in the system. Let us formulate main results that hold under some not restrictive and natural conditions.

Theorem 1 The process $\mathrm{Q}(t)$ is stochastically bounded iff $\rho < 1$, where

$\rho = \frac{\lambda E\tau^{(2)}}{\mu}$, $\mu = \sum_{i=1}^{m} E \sum_{l=1}^{\nu_{i1}} H_i(u_{i1}^{(2)})$.

Keywords: Regeneration, Unreliable Server, Multiserver System.

Acknowledgment: *The research was partially supported by RFBR grant 17-01-00468.*

Statistical Analysis of Regression Models under Grouping of the Dependent Variable

Helena Ageeva, Yuriy Kharin
Belarusian State University, Belarus

Consider a regression model under grouping of the dependent variable, when instead of the exact value one observes only the interval the hidden value belongs to. The set of possible intervals is fixed and divides the set of real numbers into nonintersecting subsets. As the result we observe a set of independent, but not identically distributed discrete random variables that represent the numbers of the corresponding intervals. Our objective is to construct statistical inferences based on these incomplete regression data.
We present the following results:
1) Conditions, under which the MLE for the model parameters is consistent and asymptotically normally distributed.
2) Statistical tests for parametric simple and composite hypotheses.
3) A modification of chi-squared test for goodness-of-fit for two cases: when the null hypothesis is simple and when the null hypothesis is specified by a family of parametric functions.
4) Results of computer experiments.

Keywords: Regression model, Grouped data, MLE, Goodness-of-fit testing, Chi-squared test.

A Semiparametric Model in Environmental Data

Giuseppina Albano, Michele La Rocca, Cira Perna
Dept of Economics and Statistics, University of Salerno, Italy

Air quality data, such as all environmental data, are tipically characterized by some stylized facts, such as non-stationarity, non linearity, seasonality, conditional heteroscedasticity and missing data. In particular, missing values make difficult to determine whether the limits set by the European Community on certain indicators of air quality are fulfilled or not. Indeed, due to missing values, the number of exceedances per year of PM10, that is particulate matter 10 micrometers

or less in diameter, and other air quality indicators is often heavily underestimated and no environmental policy is applied to protect citizen health.
In this paper we propose a model for PM10. It combines a local estimator for the trend-cycle and a ARMA-GARCH model with exogenous components as an estimator of the detrended time series. The first choice provides a flexible local structure which are not influenced by missing values outside the estimation window. The ARMA-GARCH model is able to capture the residual non linear structure and heteroschedasticity in the data. Exogenous variables are also included in the model since data of PM10 in near stations are genrally influenced by same geographic and weather conditions.
In order to validate the model a cross validation approach is performed. An application to real data to the case of Northern Italy is also presented and discussed.

Using Deepest Dependency Paths to Enhance Life Expectancy Estimation

Irene Albarrán[1], Pablo Alonso-González[2], Ana Arribas-Gil[3], Aurea Grané[1]
[1]Statistics Dept, Universidad Carlos III de Madrid, Spain,
[2]Statistics Dept, Universidad de Alcalá, Spain, [3]Acción Contra El Hambre, Spain

Dependency, that is, lack of autonomy in performing basic activities of daily living (ADL) can be seen as a consequence of the process of gradual aging. In Europe in general and in Spain in particular this phenomena represents a problem with economic, political and social implications. The prevalence of dependency in the population, as well as its intensity and evolution over the course of a person's life are issues of greatest importance that should be addressed. From EDAD 2008 (Survey about Disabilities, Personal Autonomy and Dependency Situations, INE) Albarrán-Lozano, Alonso-González and Arribas-Gil (J R Stat Soc A 180(2): 657–677, 2017) constructed a pseudo panel that registers personal evolution of the dependency rating scale and obtained the individual dependency paths. These dependency paths help to identify different groups of individuals attending to the distances to the deepest path of each age/gender group. To estimate life expectancy free of dependency (LEFD) we consider several scenarios according to dependency degree (moderate, severe and major) and the distances to the deepest path. Via Cox regression model we obtain the 'survival' probabilities (in fact, the staying free of dependency probability at a given age given that a person is alive at that age). Then marginal probabilities are obtained by multiplying these estimates by survival probabilities

given by the Spanish disabled pensioners' mortality table. Finally, we obtain the LEFD for Spanish dependent population considering gender, dependency degrees and ages from 50 to 100.

Keywords: ADL, Cox Regression, Dependency, Disability, Functional Data.

Robust Analysis of Retrial Queues

Lounes Ameur, Louiza Berdjoudj, Karim Abbas
Dept of Mathematics, Applied Mathematics Laboratory, University of Bejaia, Algeria

In this paper, we investigate the M/M/1 retrial queue with finite size orbit, working vacation interruption and classical retrial policy, where we show how parameter uncertainty can be incorporated into the model. We will assume that the retrial rate to be a random variable of distribution which obtained from sample statistic. Therefore, we illustrate the impact of computing the performance measures of the model with an analysis of parameter uncertainty. Specifically, we will provide an approach based on Taylor series expansion for Markov chains for evaluating the uncertainty in the performance measures of the considered queuing model. We develop an algorithm for evaluating the expected value and the variance of different performance measures under the assumption that the retrial rate is computed with uncertainty. Several numerical examples are carried out to illustrate the accuracy of the proposed approach.

Keywords: Retrial queues, Taylor-series expansions, Parameter uncertainty, Robustness analysis, Algorithm.

Predictive Analytic Modelling of Clinical Trials Operation

Vladimir Anisimov
Statistical Consultant & Honorary Professor in the School of Mathematics and Statistics, University of Glasgow, United Kingdom

The complexity of large clinical trials and stochasticity and multi-state hierarchic structure of various operational processes require developing predictive analytical techniques for efficient modelling and predicting trial operation using stochastic processes with random parameters in the empirical Bayesian setting.
Forecasting patient enrolment is one of the bottleneck problem as uncertainties in enrolment substantially affect trial time completion, supply chain and associated costs. In the talk an analytic methodology

for predictive patient enrolment modelling developed by Anisimov & Fedorov (2005–2007) using a Poisson-gamma model is extended further to risk-based monitoring interim trial performance of different metrics associated with enrolment, screen failures, various events, AE, and detecting outliers.
The next stage is devoted to describing an analytic methodology for modelling event counts in event-driven trials at trial start-up and interim stages including multiple events and lost patients. Some approaches to optimal trial design accounting for enrolment and follow-up stages and to risk-based monitoring and detection unusual event patterns in centres/countries are proposed.
As the next step, to model various hierarchic processes on the top of enrolment, a new methodology using evolving stochastic processes is developed. This class provides rather general and unified framework to describe various operational processes including follow-up patients, patient's visits, various events and associated costs. The technique for evaluating predictive distributions, mean and credibility bounds for evolving processes is developed (Anisimov, 2016). Some applications to modelling operational characteristics in clinical trials are considered. For these models, predictive characteristics are derived in a closed form, thus, Monte Carlo simulation is not required.

Keywords: Clinical Trial, Enrolment, Poisson-Gamma Model, Estimation, Prediction, Bayesian Re-estimation.

Switching Processes in Queueing Models

Vladimir Anisimov
Statistical Consultant & Honorary Professor in the School of Mathematics and Statistics, University of Glasgow, United Kingdom

The talk is devoted to the asymptotic investigation of switching queueing systems and networks. The methods of analysis use limit theorems of averaging principle and diffusion approximation types for the class of "Switching Processes" invented by the author. This class can serve as a convenient tool to describe state-dependent stochastic models with random switching due to internal and external factors.
Different classes of overloaded switching queueing models (heavy traffic conditions) operating under the influence of internal and external random environment, e.g. state-dependent models with Markov and semi-Markov switching, are investigated in transient regimes. The asymptotic state-space aggregation (approximation of hierarchic models with fast switching by simpler models with aggregated state-space and averaged transition rates) is also considered.

These results form an advanced methodology for the investigation of transient phenomena for hierarchic queueing models in heavy-traffic conditions and provide an analytic approach to performance evaluation of queueing networks of a complex structure [1]. Different examples are considered.

Keywords: Queueing models, networks, switching process, averaging principle, diffusion approximation, heavy-traffic, asymptotic aggregation.

References:

[1] V. Anisimov, Switching Processes in Queueing Models, John Wiley & Sons, ISTE, London, 2008.

Identifying the Boundaries Application in the Study of HRV

Valery Antonov, Artem Zagainov
Departament of Mathematics,
Peter the Great St. Petersburg Polytechnic University, Russia

The work is devoted to revelation of possibilities from multifractal numerical procedures and similar handling techniques of time series, developed until now. Application FracLab 2.1 from the mathematic packet MatLab R2008b was used for this purpose. It processes test data (generated, for example, by the Henon map, Lorenz system of equation sand etc.), bases of heart rate variability (HRV) records from the PhysioNet server and own electrocardiograms (ECG). Three-dimensional graphics of received wavelet transformation, graphs of isolines, multifractal characteristics in the form of Holder exponent and scaling exponent were constructed. All possible data provided in the FracLab (for example, Mhat-wavelet, DoG-wavelet and etc.) were used as a wavelet-forming function in the course of calculating a spectrum. Disadvantages of used realization, received first of all for test data calculating the Holder exponent, were determined. The conception of multifractal methodology widening was proposed for research of time series and creation of own software on its basis.

The scaling or scaling invariance is a main feature which is searched in the time-series in the analysed methods of researches. The most advanced method WTMM (wavelet transform modulus maxima) uses plotting of the whole variety of the local maximum line of wavelet transform. The approach to plotting the scaling is based on analysing all lines by introducing the partial function with the weight degree of all wavelet transform maximums. Above approach has built capacity scaling, ranging from construction of the expansion of the original time series, setting the scale invariance of a certain line and ending with the introduction of the special function lines for finding local maxima of the

generalized scaling. This is caused by frequent failure results in the construction of scaling a single line.

Keywords: heart rate variability, method of wavelet transforms modulus maxima.

Medical and Demographic Consequences of the Stressful Living Conditions

Valery Antonov[1], Anatoly Kovalenko[2], Konstantin Lebedinskii[3]
[1]Dept of Mathematics, Peter the Great St. Petersburg Polytechnic University, Russia, [2]Ioffe Institute, Russian Academy of Sciences, Russia, [3]Dept of Anaesthesiology and Reanimation, St. Petersburg Medical Academy of Postgraduate Studies, Russia

The report presents an analysis of demographic and health consequences of the stressful living conditions for the progression of various diseases and mortality in different age groups, from babies to centenarians. The information was taken according to Russian and foreign statistical data of the pre-war, blockade, postwar and modern age. For today, in different countries the number of people, who survived in the blockade of Leningrad, is approximately equal to 200,000. The data on the health state of 2000 such citizens and members of their families is available in the literature. Analysis of this data indicates that the prolonged fasting and psychophysiological stress have crucial effect on the health of not only the inhabitants of the besieged city, but also their children, grandchildren and great-grandchildren.

The long-term effects of stressful living conditions clearly manifest in the people of blockade and their descendants in other stressful periods, particularly during the perestroika years. Statistics show that in this period mortality rate among people, who have reached pre-retirement and retirement age was close to the rate of blockade time.

In General, all the above have contributed to a serious decline in the genetic pool of the Leningrad citizens. Many of them were evacuated and scattered not only across the country but also around the world. Thus the remote consequences of stressful living conditions now occur even in those people who were prosperous in developed countries, for example in Canada.

Keywords: Demographic Consequences, Stressful Living Conditions.

The Half-Logistic Lomax Distribution for Lifetime Modeling

Masood Anwar, Jawaria Zahoor
Dept of Mathematics,
COMSATS Institute of Information Technology, Pakistan

In this paper, we introduce a new two-parameter lifetime distribution called the half-logistic Lomax (HLL) distribution. The proposed distribution is obtained by compounding half-logistic and Lomax distributions. We obtain some mathematical properties of the proposed distribution such as the survival and hazard rate function, quantile function, mode, median, moments and moment generating functions, mean deviations from mean and median, mean residual life function, order statistics and entropies. The estimation of parameters is performed by maximum likelihood and provide formulas for the elements of the Fisher information matrix. A simulation study is provided to access the performance of maximum likelihood estimators (MLEs). The flexibility and potentiality of the proposed model is illustrated by means of a real data set.

Keywords: Half-logistic distribution, Lomax distribution, Mean residual life.

The Problem of the SARIMA Model Selection for the Forecasting Purpose

Josef Arlt, Peter Trcka, Markéta Arltová
Dept of Statistics and Probability,
University of Economics, Czech Republic

The goal of the work is the study of the ability to choose the proper models for the time series generated by SARIMA processes with different parameter values and to analyze the accuracy of the forecasts based on the selected models. The work is based on the simulation study. For this purpose a new automatic SARIMA modelling method is proposed and used. Also the other competing automatic SARIMA modelling procedures are applied and the results are compared. The important question to which the reference should be made is the relation of the magnitude of the SARIMA process parameters i. e. the size of the systematic part of the process and the ability to select a proper model. Another addressed problem is the relationship between the quality of the selected model and the accuracy of forecasts achieved by its application. The simulation study leads to the results that can be generalized to most empirical analyses in various research areas.

Keywords: Time series, modelling, SARIMA, simulation.

Acknowledgements: *This paper was written with the support of the Czech Science Foundation Project No. P402/12/G097 DYME - Dynamic Models in Economics.*

The Extended Flexible Dirichlet Model: a Simulation Study

Roberto Ascari[1], Sonia Migliorati[2], Andrea Ongaro[2]

[1]Department of Statistics and Quantitative Methods (DISMEQ), Università di Milano-Bicocca, Italy, [2]Department of Economics, Management and Statistics (DEMS), Università di Milano-Bicocca, Italy

Compositional data are prevalent in many fields (e.g. environmetrics, economics, biology, etc.). They are composed by positive vectors subject to a unitsum constraint (i.e. they are defined on the simplex), proportions being an example of this kind of data. A very common distribution on the simplex is the Dirichlet, but its poor parametrization and its inability to model many dependence concepts make it unsatisfactory for modeling compositional data. A feasible alternative to the Dirichlet distribution is the Flexible Dirichlet (FD), introduced by Ongaro and Migliorati [1]. The FD is a generalization of the Dirichlet that enables considerable flexibility in modeling dependence as well as various independence concepts, though retaining many good mathematical properties of the Dirichlet. More recently, the Extended Flexible Dirichlet (EFD, [2]) distribution has been proposed in order to generalize the FD. The EFD preserves a finite mixture structure as the FD, but it exhibits some relevant advantages over the FD, such as a more flexible cluster structure and a (even strong) positive dependence for some pairs of variables. The aim of this contribution is twofold. First we propose and investigate sophisticated EM algorithms for parameters estimation, with particular emphasis on the initialization problem, which is a crucial issue. Furthermore, we devise a simulation study to evaluate the performances of the MLE of the parameters as well as of a procedure proposed to compute their standard errors.

Keywords: Compositional Data, Dirichlet Mixture, EM algorithm.

References:

[1]. A. Ongaro and S. Migliorati. A generalization of the Dirichlet distribution. Journal of Multivariate Analysis, 114, 412{426, 2013.

[2]. A. Ongaro and S. Migliorati. A Dirichlet mixture model for compositions allowing for dependence on the size. In M. Carpita, E. Brentari and E.M. Qannari (Eds), Advances in Latent Variables Methods, Models and Applications, Springer, 2014.

[3]. S. Migliorati and A. Ongaro. Estimation issues in the Extended Flexible Dirichlet model. 16th ASMDA Conference Proceedings, Greece, 2015.

Sojourn Time and Busy Period in a Discrete-Time Queueing System with Breakdowns

Ivan Atencia
Dept of Applied Mathematics, University of Málaga, Spain

In this paper it is analyzed one of the most import scenes in a queueing systems, that is, the time that a customer spends in the server and in the system, as-well as the period in which a customer arrives finding the system empty until it leaves the system with no customers behind him. Customers are served with a discrete general distribution under a Last Come First Serve Discipline (LCFS). The possibility of expulsions out of the system and breakdowns while the server in functioning is also contemplated. It is applied the mathematical tool of generating functions in order to find the main performance characteristics of the model.

Keywords: discrete-time, sojourn time, busy period, Generating Functions.

Estimates for Initializing Enrollment Models

Matthew Austin
Development Design Center, Amgen Inc, USA

Historic performance data forms the basis for initial enrollment modeling of future studies. Sponsors of clinical studies generally use investigator site performance from previous studies to derive estimates for modeling new studies. Traditionally, sponsors use marginal subgroup estimates of performance data where the subgroup is formed based on design characteristics such as disease, study phase, and country. New studies may share some characteristics with previous studies; however, it is common for studies within the same phase and indication to enroll subjects in different countries. Using hierarchical modeling allows borrowing of information for new countries from other indications or phases where these countries participated in previous studies.
The positives and negatives of the hierarchical modeling approach in this setting will be discussed as well as prediction when conditioning on a subset of variables in the multi-level negative binomial model case.

Keywords: Hierarchical models, multi-level models, mixed-effect models, negative binomial, estimation.

Numerical Stability of the Escalator Boxcar Train under reducing System of Ordinary Differential Equations

Tin Nwe Aye, Linus Carlsson, Sergei Silvestrov
Division of Applied Mathematics, Mälardalen University, Sweden

The Escalator Boxcar Train (EBT) is one of the most popular numerical methods to study the dynamics of physiologically structured population models. The EBT-model can be adapted to numerically solve population dynamics of ecological and biological systems with continuous reproduction. The original EBT-model accumulates a dynamic system of ODE's to solve in each time step.
In this project, we propose an EBT-solver to overcome some computational disadvantageous of the EBT method which includes the automatic feature of merging and splitting the cohorts, in particular we apply the model to a colony of Daphnia pulex.

Keywords: escalator boxcar train, physiologically structured population models, daphnia model.

Real-time Monitoring and Control of Industrial Processes using Electrical Tomography Data

Robert G Aykroyd
Dept of Statistics, School of Mathematics,
University of Leeds, United Kingdom

Electrical tomography methods offer the potential for cheap and non-invasive monitoring and control of dynamic processes in industry and medicine. Although image reconstruction is useful for process visualization, automatic control does not require an image. For process monitoring, estimation of key geometric and process-control parameters is more appropriate than visualization. Such parameters can then be used as feedback for process adjustment, avoiding the need for human intervention. To demonstrate the proposed framework, a simple simulation study is described in which near real-time tracking of an object using electrical tomographic data is considered. The boundary-element method is used, in preference to the more usual finite-element method for the numerical solution of the governing physical equation, as it better matches the proposed geometrical modelling and provides more rapid calculations. The motivating process-control example considered is that of water/oil separation in a hydrocyclone.

Keywords: Geometrical models, industrial process control, inverse problems, maximum likelihood estimation, object tracking.

How to Compute the Rise Time of the acquisition of Consonants

Elena Babatsouli
Institute of Monolingual and Bilingual Speech, Greece

Determining the time a child takes to produce consonants in an adult-like manner, from occasional accuracy to repeated accuracy, is of interest in the area of child speech development. This length of time is called here the rise time of the acquisition of consonants. It is well known that consonants are acquired at different ages but it is not known whether the rise time of their acquisition is different. Rise times can be compared once the low and high accuracy levels are defined. These levels are set equal to 10% and 90%, respectively, the proportions of accurate productions to the number of attempts. When the attempted consonants are not produced accurately, they can be identified from the rest of the produced word. The difficulty in computing this rise time lies in having dense longitudinal data in order to capture the true rise time, that is, the difference between the earliest age of 90% accuracy level and the latest age of 10% accuracy level. In this study, data are employed from a child's daily word productions in English during speech interactions with the author, from age two and a half years to age four years. The data were digitally recorded and subsequently phonetically transcribed in a CLAN database by the author. The accuracy level of produced consonants was averaged weekly. It is found that all consonants are acquired at the 90% accuracy level by the age of three years and eleven months, with *r* being acquired last, having the shortest rise time of three months. The consonants acquired earlier, *h, θ, ð, k, g*, have a rise time which is two months longer. *ð* is acquired at the age of three years and seven months while *h, θ, k, g* are acquired three months later. It is seen that the age of a consonant's acquisition, besides the degree of articulation difficulty, also depends on the complexity of the words in the child's vocabulary that contain the consonant. This is attested by the acquisition of *ð* earlier than *θ,* even though *ð* is harder to articulate. Furthermore, there seems to be a critical age, call it phonological age of maturity, for this child being around three years and eight months, after which the accuracy of consonants rises fast. It will be interesting to apply the proposed methodology to other children and see whether similar conclusions can be drawn concerning the rise time of the acquisition of consonants in general.

Truncation of Markov Chains with Applications to Queueing

Badredine Issaadi, Karim Abbas, Djamil Aïssani
Laboratory LAMOS, University of Bejaia, Algeria

The calculation of the stationary distribution for a stochastic infinite matrix is generally difficult and do not have closed form solutions, it is desirable to have simple approximations converging rapidly to this distribution. In this paper, we use the weak perturbation theory to establish analytic error bounds in the GI/M/1 model and a tandem queue with blocking. Numerical examples are carried out to illustrate the quality of the obtained error bounds.

Keywords: Truncation, Queuing System, weak Stability, Simulation, Algorithm.

ALT Modeling when the AFT Model Fails

Vilijandas Bagdonavičius, Rūta Levulienė
Dept of Mathematical statistics, Faculty of mathematics and informatics, Vilnius University, Lithuania

The accelerated failure time (AFT) model is the most used model in accelerated life testing (ALT). This model is restrictive because the failure time distributions under different constant stresses differ only in scale. If it is not so, most of papers on ALT use a generalization of the AFT model supposing that under different stresses not only scale but also shape parameters are different. This model has one very unnatural property: the hazard rates (and the survival functions) corresponding to usual and accelerated stresses intersect.
We consider a flexible generalization of the AFT model including not only crossing of hazard rates but also wide class of alternatives of non-intersecting hazard rates which may approach, go away, be proportional, and/or coincide.
Estimation procedures and properties of estimators are discussed. Examples of data analysis are given. Goodness-of-fit tests for the given models are proposed

Keywords: Accelerated life testing, estimation, accelerated failure time, goodness-of-fit, stress.

Some Issues in Predicting Patient Recruitment in Multi-Centre Clinical Trials

Andisheh Bakhshi[1], Stephen Senn[2], Alan Phillips[3]
[1]*Institute of Health and Wellbeing, Robertson Centre for Biostatistics, University of Glasgow, United Kingdom,*
[2]*Competence Center for Methodology and Statistics, CRP-Santé, Luxembourg,*
[3]*ICON Clinical Research UK Ltd, United Kingdom*

A key paper in modelling patient recruitment in multi-centre clinical trials is that of Anisimov and Fedorov. They assume that the distribution of the number of patients in a given centre in a completed trial follows a Poisson distribution. In a second stage, the unknown parameter is assumed to come from a Gamma distribution. As is well known, the overall Gamma-Poisson mixture is a negative binomial. For forecasting time to completion, however, it is not the frequency domain that is important, but the time domain and that of Anisimov and Fedorov have also illustrated clearly the links between the two and the way in which a negative binomial in one corresponds to a type VI Pearson distribution in the other. They have also shown how one may use this to forecast time to completion in a trial in progress. However, it is not just necessary to forecast time to completion for trials in progress but also for trials that have yet to start. This suggests that what would be useful would be to add a higher level of the hierarchy: over all trials. We present one possible approach to doing this using an orthogonal parameterization of the Gamma distribution with parameters on the real line. The two parameters are modelled separately. This is illustrated using data from 18 trials. We make suggestions as to how this method could be applied in practice.

Keywords: negative binomial, clinical trials, recruitment.

Clustering of Spatially Dependent Data Streams based on Histogram Summarization

Antonio Balzanella, Rosanna Verde, Antonio Irpino
Università della Campania "Luigi Vanvitelli", Italy

In the framework of data stream analysis, we introduce a new strategy for clustering data sequences recorded, on-line, by spatially located sensors and for monitoring their evolution over the time. Our proposal is based on a first summarization of the sub-sequences in non-overlapping windows, through histograms. Then, a three-phase strategy is performed. The first phase is the partitioning of the sub-sequences in each window into clusters, considering the spatial dependence among

the sensors data; the second phase is the updating of a suitable dissimilarity matrix; finally, the third step performs a final clustering analysis on the dissimilarity matrix of the local partitions in order to obtain a final partition of the streams. The data streams change is detected by considering the evolution of the local partitions over a user chosen time period and the evolutions in proximity relations. The comparison between histograms and the incorporation of the spatial dependence in the clustering process is performed using a weighted L2 Wasserstein metric. Through applications on real and simulated data, we show this method provides results comparable to algorithms for stored data.

Keywords: Data stream, Clustering, Spatial data.

Ordinal Regression with Geometrical Objects Predictors: An Application to Predict the Garment Size of a Child

Sonia Barahona, Pablo Centella, Ximo Gual-Arnau, María Victoria Ibañez, Amelia Simó
Dept of Mathematics, Universitat Jaume I, Spain

The aim of this work is to model an ordinal response variable in terms of a functional predictor included on a vector-valued Reproducing Kernel Hilbert Space (RKHS). Our modelization is based on functional regression [1], using as orthonormal base the eigen functions of the integral operator defined by the kernel. This is an alternative to the popular principal component functional regression approach.
In particular, we focus on the vector-valued RKHS obtained when the contour of a geometrical object (body) is characterized by a Current [2]. This approach is applied to predict the fit of a garment size to a child, based on a 3D scan of his/her body. Size fitting is a significant problem for online garment shops. Our data was obtained from a fit assessment study carried out on a sample of the Spanish child population. Children were measured using a 3D body scanner and different garments of different sizes were tested on them. An expert assessed their fit and classified it in terms of Large, Correct or Small.

Keywords: Currents, Statistical Shape and Size Analysis, Reproducing Kernel Hilbert Space, Functional Regression.

References:

[1] B. Silverman, and J. Ramsay. *Functional Data Analysis*. Springer, 2005.
[2] S. Durrleman, X. Pennec, A. Trouvé, and N. Ayache. Statistical models of sets of curves and surfaces based on currents. *Medical image analysis*, 13(5):793 808, 2009.

A Bivariate Mixed-Type Distribution with Applications to Reliability

Alessandro Barbiero
Dept of Economics, Management and Quantitative Methods, Università degli Studi di Milano, Italy

Mixed continuous and discrete data are commonly seen nowadays in many scientific fields, for example in health and behavioral sciences. Although several proposals have been suggested, the stochastic modeling of joint mixed-type distributions is still challenging.
Resorting to Farlie-Gumbel-Morgenstern copula, a bivariate mixed-type distribution, with geometric and exponential components, is proposed. Its properties are explored, with a particular focus on range of correlations (it is a well-known fact that Farlie-Gumbel-Morgenstern copula allows for a moderate level of correlation) and reliability concepts. Parameter estimation is examined, with a special attention at the parameter accommodating the dependence structure; estimation techniques are proposed and assessed also via Monte Carlo simulation experiments. An application to real data is provided.

Keywords: bivariate models, copula, exponential distribution, geometric distribution, reliability.

Step semi-Markov Models and Application to Manpower Management

Vlad Stefan Barbu[1], Guglielmo D'Amico[2], Raimondo Manca[3], Filippo Petroni[4]
[1]Université de Rouen, Laboratoire de Mathématiques Raphael Salem, France, [2]Dipartimento di Farmacia, Universita "G. d'Annunzio" di Chieti-Pescara, Italy, [3]Dept of Methods and Models for Economics, Territory and Finance, Universita "La Sapienza" Roma, Italy, [4]Dipartimento di Scienze Economiche ed Aziendali, Universita di Cagliari, Italy

The purpose of this paper is to introduce a class of stochastic processes that we call step semi-Markov processes and to illustrate the modeling capacity of such processes in practical applications. The name of this process comes from the fact that we have a semi-Markov process and the transition between two states is done through several steps. We first introduce these models and the main quantities that characterize them. Then, we derive the recursive evolution equations for two-step semi-Markov processes. The interest of using this type of model is illustrated by means of an application in manpower planning.

Keywords: semi-Markov processes, manpower management, stochastic modeling.

Human Factors: Coal Face Reality of Recruitment to Clinical Trials

Katharine D Barnard
Bournemouth University, BHR Limited, United Kingdom

Fewer than a third of publicly funded trials recruit according to original plans, prolonging the duration of the trial, increasing costs and delaying delivery of answers to important research questions that could benefit the lives of patients.
Professor Barnard's talk will highlight some of the human factors associated with failure to recruit to clinical trials and what we can do about it. Under pressure to minimise costs, optimize outcomes and 'outbid' competitor research groups to secure funding, human factors often go unnoticed however they are potentially devastating to delivery of the trial. She will explore the costs associated with failure to recruit as planned, demonstrating that financial costs are obvious, however presenting arguments that other costs are also impactful and deleterious such as failure to answer key research questions effectively, failure to further research, failure to adequately inform health policy nor deliver interventions that may be clinically and cost-effective for patients and improve their quality of life. Accurate estimation of sample size to adequately power a trial is crucial and undisputed, however equally crucial is delivering the research and providing answers to clinically relevant, meaningful and important questions. Ensuring the human factors associated with recruitment of participants and centers is both recognized and addressed will help improve recruitment and delivery of high quality trials.

Keywords: Clinical trials recruitment, human factors, barriers, facilitators.

Diffusions and Generalised Logistic Dynamics

Antonio Barrera-García[1], Patricia Román-Román[2], Francisco Torres-Ruiz[2]
[1]Departamento de Matemática Aplicada, E.T.S.I. Informática, Universidad de Málaga, Spain, [2]Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Granada, Spain

Stochastic diffusion processes have become an appropriate tool to explain dynamical phenomena, taking into account random influence

produced by internal and external conditions. Common approaches are based on stochastic extensions of deterministic classical models. One of the most important is the continuous version of the logistic growth curve, widely used to describe sigmoidal growth. This success has produced several models based on logistic curve, extending the latter by more sophisticated exponential functions. These specific logistic-based models focus their improved performance on different aspects. For instance, type I hyperbolastic growth curves have been successfully applied to stem cell growth or epidemiological dynamics, due to their flexible condition. Nevertheless, the increasing number of models, in addition to their sophistication, can restrict the applications. For this reason, a generalized viewpoint, determining a rigorous way of mathematical abstraction, is required. We aim to establish theoretical framework to define a logistic-type diffusion process based on a functional generalization of the classical, deterministic logistic model. This leads to a characterization of several models as particular cases of a general theory, as well as introducing some mathematical developments which can help to determine new perspectives of logistic growth dynamics.

Keywords: Diffusion process, growth curve, logistic model, functional generalization.

Limit Theorems for Infinity Chanel Queueing Systems with a Regenerative Input Flow

Elena Bashtova, Ekaterina Chernavskaya
Dept of Probability, Faculty of Mathematics and Mechanics, Lomonosov Moscow State University, Russia

We consider an infinity channel queueing system S with a regenerative input flow. The main feature of the system is that the service times are so heavy tailed that the average service time is infinite. Our main purpose is the asymptotic analysis of the process q(t) which is equal to the number of occupied servers in system S at time t. Absence of average service time leads to fact that the number of customers in the system tends to infinity over time. We prove analogues of the Law of Large Numbers and Central Limit Theorem for the process q(t). Our proofs are based on majorization method, estimations of rate convergence in the classical LLN and some inequalities concerning demimartingales.

Keywords: Limit Theorems, Regenerative Input Flow, Heavy Tails, Demimartingales.

Acknowledgment: *The research was partially supported by RFBR grant 17-01-00468.*

Predicting of Least Limiting Water Range (LLWR) of Soil using MSECE Model

Behnam Behtari, Adel Dabbag, Mohammadi Nasab
Dep. Crop Ecology, Faculty of Agriculture,
Uinversity of Tabriz, East Azerbaijan, Iran

The least limiting water range, LLWR, is the range of soil water content within which plant growth is least limited by water potential, aeration, and mechanical resistance and outside of this range, Limited access to water for plant was increased. The aim of this work is to offer a simple application called MSECE V.1.0 in Microsoft Excel software, to predict LLWR of soil. Inputs data requirement to calculation matric potential of soil were soil components percentage of sand, silt and clay, volumetric water content, saturated volumetric water content and soil bulk density. Soil matric potential value calculated by the application together with soil organic matter and soil bulk density were input requirement to obtained LLWR and then calculate the amount of available water in the soil. Soil samples were taken from field experimental of beans and corn in three regions of Ahar, Tabriz and Ardebil in the years 2010, 2014 and 2015, respectively. The a and b parameters in water release curve were effect of bulk density and matric suction on moisture retention, respectively and c and e parameters in Soil resistance curve were effect of moisture and bulk density on soil mechanical resistance, respectively.

Keywords: Available water content, Simulation, Modeling and MSECE software.

Quantifying the Sensitivity of Bush Bean and Maize Seed Germination to Soil Oxygen using an Oxygen-Time Threshold Model

Behnam Behtari, Adel Dabbag, Mohammadi Nasab
Dept Crop Ecology, Faculty of Agriculture, Uinversity of Tabriz, Iran

Soil oxygen is one of the main requirements for germination. The O_2 requirement for seed germination is also strongly modulated by other environmental factors. The field experiments were carried out in randomized complete block design in three regions of Ahar, Tabriz and Ardabil in 2010, 2014 and 2015 respectively to determine of sensitivity of bush bean and maize seed germination to soil oxygen. Daily seed germination percentage and soil oxygen for each crop were predicted by using the MSECE model. In all of the experiments, the percentage of soil oxygen was significant at 1% probability levels. The highest soil oxygen average 4.4% was belonged to Ardabil with silt loam soil. Oxygen

threshold model showed that the germination time can be reduced to 50% its value when the percentage of oxygen respectively for beans and corn to 4 and 3.4 percent. The highest frequency was determined 5.79% for oxygen, which probably happened in the days when the soil water content is lower. In this study, it was found that this model could be used to quantifying the response of crops to soil oxygen.

Accounting for Model Uncertainty in Mortality Projection

Andrés Benchimol[1], J. Miguel Marin[1], Irene Albarrán[1], Pablo Alonso-González[2]
[1]Statistics Dept, Universidad Carlos III de Madrid, Spain,
[2]Economics Dept, Universidad de Alcalá, Spain

Forecasting mortality rates has become a key task for all who are concerned with payments for non-active people, such as Social Security or life insurance firms. The non-ending process of reduction in the mortality rates is forcing to improve continuously the models used to project these variables. Traditionally, actuaries have selected just one model, supposing that this specification generated the observed data.
Most of the times results have driven to a set of questionable decisions linked to those projections. This way to act does not consider the model uncertainty when selecting a specific one. This drawback can be reduced through model assembling. This technique is based on using the results of a set of models in order to get better results.
In this paper we introduce two approaches to ensemble models: a classical one, based on the Akaike information criterion (AIC), and a Bayesian model averaging method.
The data are referred to a Spanish male population and they have been obtained from the Human Mortality Database. We have used four of the most extended models to forecast mortality rates (Lee-Carter, Renshaw-Haberman, Cairns-Blake-Dowd and its generalization for including cohort effects) together with their respective Bayesian specifications. The results suggest that using assembling models techniques gets more accurate projections than those with the individual models.

Keywords: AIC model averaging, Bayesian model averaging, bootstrap, Cairns-Blake-Dowd model, Lee-Carter model, longevity risk, model uncertainty, projected life tables, Renshaw-Haberman model.

Measuring Latent Variables in Space and/or Time. A Latent Markov Model Approach

Gaia Bertarelli[1], Franca Crippa[2], Fulvia Mecatti[3]
[1]*Dept of Economics, University of Perugia Via Alessandro, Italy*
[2]*Dept of Psychology, University of Milano-Bicocca, Italy,*
[3]*Dept of Sociology, University of Milano-Bicocca, Italy*

Composite indicators have the advantage of syntesising a latent, multi-dimensional dimension in a single digit, usually included in the interval (0; 1). They can be computed as a weighted sum of indicators, or as a measure of a latent variable, as in Structural equation Models (SEM), traditionally applied to obtain a single measure. When the latent variable is considered to have a dynamic of its own, a new outlook in the construction of composite indicators is provided by Multivariate Latent Markov model (LMM). LMMs are a particular class of statistical models which assume the existence of a latent process affecting the distribution of the response variables.

The main assumption is conditional independence of the response variables given the latent process, which follow a first order dicrete Markov chain with a finite number of states. The basic idea underlying the approach is that the latent process fully explains the observable behaviour of an item together with available covariates.

Analogously to SEM, LMMs are composed of two parts: a measurement model, concerning the conditional distribution of the response variables given the latent process, and the latent model, pertaining the distribution of the latent process. LMMs can account for measurement errors or unobserved heterogeneity between areas in the analysis. LMMs main advantage is that the unobservable variable is allowed to have its own dynamics and it is not constrained to be time constant. In addition, when the latent states are identified as different subpopulations, LMMs can indentify a latent clustering of the population of interest, with areas in the same subpopulation having a common distribution for the response variables. Under this respect, a LMM may be seen as an extension of the latent class (LC) model, in which areas are allowed to mode between the latent classes during the observational period. Available covariates are included in the latent model and then may affect the initial and transition probabilities of the Markov chain.

Our applicative viewpoint intends to adapt the LMM approach to a synthetic index.

In our case, we focus on gender gap as the latent status - both in space and time. The gap is in fact a latent trait, namely only indirectly measurable through a collection of observable variables and indicators purposively selected as micro-aspects contributing to the latent macro-dimension.

Keywords: latent, markov chain, index, mixture model.

Semi-Parametric Consistent Estimators for Recurrent Event Times Models based on Parametric Virtual Age Functions

Eric Beutner[1], Laurent Bordes[2], Laurent Doyen[3]
[1]Dept of Quantitative Economics, Maastricht University, The Netherlands, [2]Univ. Pau & Pays Adour, Laboratoire de Mathématiques et de leurs Applications, France, [3]Univ. Grenoble Alpes, France

We consider a large class of semi-parametric models for recurrent events based on virtual ages. Modeling recurrent events lifetime data using virtual age models has a long history. This rich class of model contains standard model families as non-homogeneous Poisson processes and renewal processes and may include covariates or random effects (see for instance Pena (2006, *Statistical Science*) for a large overview on these models). In many non- or semi-parametric works the virtual age function is supposed to be known, this weakness can be overcome by parameterizing the virtual age function (see for instance Doyen and Gaudoin, 2004, *Reliability Engineering and System Safety*). Then the model consists of an unknown hazard rate function, the infinite-dimensional parameter of the model, and a parametrically specified virtual age (or effective) function. Recently Beutner, Bordes and Doyen (2016, *Bernoulli*) derived conditions on the family of effective age functions under which the profile likelihood inference method for the finite-dimensional parameter of the model leads to inconsistent estimates. Here we show how to overcome the failure of the profile likelihood method by smoothing the pseudo-estimator of the infinite-dimensional parameter of the model, by adapting a method proposed by Zeng and Lin (2007, *Journal of the American Statistical Association*) for the accelerated failure time model.

Keywords: Recurrent events, Virtual age, Semi-parametric, Consistency.

Designing Critical Infrastructure Network with Cascading for Predefined Safety Level

Agnieszka Blokus-Roszkowska, Krzysztof Kołowrocki
Dept of Mathematics, Gdynia Maritime University, Poland

The multistate approach to cascading effect modeling in critical infrastructure (CI) networks is proposed. Describing cascading effects in

CI networks both the dependencies between subnetworks of this network and between their assets are considered. Then, after changing the safety state subset by some of assets or subnetworks to the worse safety state subset, the lifetimes of remaining assets, respectively subnetworks, in the safety state subsets decrease. Models of dependency and behavior of components can differ depending on the structural and material properties of the network, operational conditions and many other factors, as for example natural hazards. According to the equal load sharing rule, after changing the safety state subset by some of assets in the subnetwork to the worse safety state subset, the lifetimes of remaining assets in this subnetwork in the safety state subsets decrease equally depending, inter alia, on the number of these assets that have left the safety state subset. In the local load sharing model of dependency, after departure from the safety state subset by one of assets in the subnetwork the safety parameters of remaining assets are changing dependently of the coefficients of the network load growth. These coefficients are concerned with the distance from the asset that has got out of the safety state subset and can be interpreted in the metric sense as well as in the sense of relationships in the functioning of the network. Apart from the dependency of assets' departures from the safety states subsets, the dependencies between subnetworks are also taken into account.

Proposed theoretical models are applied to the safety analysis of the exemplary network of transmission lines regarding dependencies of its lines and subnetworks. For presented exemplary network, the intensities of departure from the safety state subsets of transmission lines are estimated for arbitrarily assumed values of the network mean lifetimes. Further, the estimation of intensities for predefined safety level can be helpful in designing of CI networks.

Keywords: multistate approach, ageing network, cascading effects, dependency model.

Acknowledgment: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Analysis of the Crude Oil Transfer Process and its Safety

Agnieszka Blokus-Roszkowska[1], Bożena Kwiatuszewska-Sarnecka[1], Paweł Wolny[2]
[1]Dept of Mathematics, Gdynia Maritime University, Poland,
[2]Naftoport Ltd, Poland

In the Baltic Sea region, there are many oil terminals, which perform transshipment of crude oil and refined petroleum products. Oil terminals are a key element of the petroleum supply logistics of crude oil to refineries and oil transit. The accident in the oil terminal during unloading/loading of tankers may have a long or short-term consequences for the work of the terminal, that may be associated with the socioeconomic losses and environmental costs consequences.
Considering the operation process of oil port terminal the paper focuses on processes related to the cargo movement inside the pipeline system. Technical parameters during all stages of crude oil transfer process are described. Processes of crude oil loading, discharging and internal recirculation are described and their statistical identification are given. Analyzing the crude oil transfer process and its influence on the oil port terminal and operating environment safety, potential threats of oil spill during oil transfer are identified. The accidental events that can cause oil spill in the terminal are in the paper classified with distinction of internal and external as well as root and contributing causes.
One of important causes of oil spill, is pressure upsurge inside the pipelines as a hydraulic hammer's consequence. These pressure surges can be generated by anything that causes the liquid velocity in a line to change quickly e.g., valve closure, pump trip, Emergency Shut Down (ESD) closure occurs and subsequently packing pressure. The particular attention is paid to the pressure upsurge inside the pipelines caused by sudden valve closure on the oil reloading installation in port terminal.
Finally, the discussion on protection of marine facilities against hydraulic transient pressure surges that can occur during crude transfer is performed. Some recommendations, including safety culture recommendation, are given. In this scope, training on recognizing and handling abnormal situations during oil transfer, as one of methods to prevent such kind of accidents, is proposed.

Keywords: oil transfer, operation process, pressure upsurge, oil spill.

Acknowledgment: *Paper in the scope of HAZARD project funded from the Interreg Baltic Sea Region Programme 2014-2020 under grant Agreement No #R023.*

Statistical identification of Critical Infrastructure Accident Consequences Process Part 1: Process of Initiating Events

Magda Bogalecka[1], Krzysztof Kołowrocki[2]
[1]*Dept of Industrial Commodity Science and Chemistry, Gdynia Maritime University, Poland,*
[2]*Dept of Mathematics, Gdynia Maritime University, Poland*

Some kinds of critical infrastructure accidents concerned with its safety level decrease may occur during its operation. Those accidents may bring some dangerous consequences for the environment and have disastrous influence on the human health and activity. Each critical infrastructure accident can generate the initiating event causing dangerous situations in the critical infrastructure operating surroundings. The process of those initiating events can result in this environment threats and lead to the environment dangerous degradations. To involve the interactions between the initiating events, the environment threats and environment degradation effects, semi-Markov general model of a critical infrastructure accident consequences was built.
In this part of the paper, the statistical methods such as the method of moments, the maximum likelihood method and the chi-square goodness-of-fit test are applied to the identification of the process of initiating events generated either by the critical infrastructure accident or by the loss of required safety critical level are used on the basis of statistical data coming from this process realizations.

Keywords: critical infrastructure, sea accident, potential consequences, initiating events.

Acknowledgment: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Statistical Identification of Critical Infrastructure Accident Consequences Process Part 2: Process of Environment Threats

Magda Bogalecka[1], Krzysztof Kołowrocki[2]
[1]*Dept of Industrial Commodity Science and Chemistry, Gdynia Maritime University, Poland,*
[2]*Dept of Mathematics, Gdynia Maritime University, Poland*

The risk analysis of chemical spills at sea and their consequences is based on the general model of mutual interactions between three processes: the process of the sea accident initiating events, the process

of the sea environment threats and the process of the sea environment degradation.
This paper is concerned with the identification of the second one of those three processes. The statistical identification of the unknown parameters of the process of environment degradation i.e. estimating the probabilities of this process of staying at the states at the initial moment, the probabilities of this process transitions between its states and the parameters and forms of the distributions fixed for the description of this process conditional sojourn times at their states are performed in the similar way to that presented in the part 1 of the paper.

Keywords: critical infrastructure, sea accident, potential consequences, environment threats.

Acknowledgment: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Statistical Identification of Critical Infrastructure Accident Consequences Process Part 3: Process of Environment Degradation

Magda Bogalecka[1], Krzysztof Kołowrocki[2]
[1]Dept of Industrial Commodity Science and Chemistry, Gdynia Maritime University, Poland,
[2]Dept of Mathematics, Gdynia Maritime University, Poland

The probabilistic general model of critical infrastructure accident consequences includes the process of initiating events, the process of environment threats and the process of environment degradation models.
This paper is concerned with the identification of the third one of those three processes. The statistical identification of the unknown parameters of the process of environment degradation are performed in the similar way to that presented in the part 1 and 2 of the paper.
The results of the parts 1-3 of the paper will be used in the prediction of the process of initiating events, the prediction of the process of environment threats, and the prediction of the process of environment degradation as well as in the prediction of the entire process of the critical infrastructure accident consequences.

Keywords: critical infrastructure, sea accident, potential consequences, environment degradation.

Acknowledgment: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Highly Dimensional Classification using Tukey Depth and Bagdistance

Milica Bogicevic, Milan Merkle
University of Belgrade, Faculty of Electrical Engineering, Serbia

In a recent paper "ABCDepth: efficient algorithm for Tukey depth" (arXiv, 2016), we presented the new approximate algorithm for evaluation of Tukey median and Tukey depth of a given point. The algorithm has an advantage over others due to linear complexity in high dimensions. In this work we use this algorithm combined with the notion of bagdistance introduced by Hubert, Rousseeuw and Segaert (Adv. Data Anal. Classif., 2016) in very high dimensional data sets. Several examples in classification and outlier detection will be presented and discussed. We will examine performances of our algorithm in very high dimensions which are still not achievable with existing exact algorithms.

Keywords: Big data, Tukey depth, Classification, ABCDepth.

Formulation of the Mean Squared Error for Logistic Regression. An Application with Credit Risk Data

Eva Boj del Val, Teresa Costa Cor
Universitat de Barcelona, Spain

It is showed an expression of the mean squared error of prediction for new observations when using logistic regression. The mean squared error can be expressed as the sum of the process variance and of the estimation variance. The estimation variance can be estimated by applying the delta method and/or by using bootstrap methodology. When using bootstrap, e.g. bootstrap residuals, one is able to obtain an estimation of the distribution of each predicted value. To help us in the knowledge of the randomness of the new predicted values, confidence intervals can be calculated by taking into account the bootstrapped distributions of the predicted new values. The calculus and usefulness of the mean squared error are illustrated to solve the problem of Credit Scoring. They are analyzed two sets of real credit risk data for which the probabilities of default are estimated. Other measures as are: error rates based on counting mispredictions; sensitivity, specificity, ROC curves and the Brier's score are calculated for comparison with the proposed mean squared error measures.

Keywords: Logistic regression, mean squared error, estimation variance, delta method, bootstrapping residuals, credit risk.

Prediction for Regularized Clusterwise Multiblock Regression

Stéphanie Bougeard[1], Niang Ndèye[2], Gilbert Saporta[2]
[1]*Dept of Epidemiology, Anses, France,*
[2]*Dept of Mathematics and Statistics, CEDRIC-CNAM, France*

Multiblock methods integrate the information in several blocks of explanatory variables to explain a set of dependent variables. However in many applications, multiblock techniques are used when the observations come from a homogeneous population but it often happens that the observations originate from different ones. A standard approach to obtain clusters within a regression framework is clusterwise, a.k.a. typological, regression. These methods assume that there is an underlying unknown group structure of the observations and that each cluster can be revealed by the fit of a specific regression model. In a more formal way, clusterwise regression simultaneously looks for a partition of the observations into clusters and minimizes the sum of squared errors computed over all the clusters.
We propose to combine a regularized multiblock regression with a clusterwise approach. We focus on prediction as a matter of utmost importance however not addressed in the clusterwise framework. In practice, clusterwise prediction can be used for two major goals: (i) the prediction of new observations and (ii) the selection of the unknown parameters of the clusterwise multiblock regression, i.e. the number of clusters, the number of number of components to be included in the model and the regularization parameter value, while minimizing the cross-validated prediction error. For this purpose, several original multiblock supervised classifications are checked in the field of clusterwise analysis. A simulation study is carried out to assess the prediction method performances and an empirical application is provided to illustrate the method usefulness.

Keywords: multiblock analysis, clusterwise regression, multiblock discriminant analysis, supervised multiblock classification.

Measuring the Shape of Voting Districts with Unchangeable Boundary

James R. Bozeman, Ryleigh E. Costigan, Abby Salvadore
Lyndon State College, USA

In this talk we show how to measure the shape of a legislative district in which some of its boundary is unchangeable. This is accomplished by using the 'convexity ratio' and the convex hull (exogon) as usual, but

finding the endogon after cutting off any intrusions into the district itself formed by immovable boundaries (caused by other states, e.g.). We conclude by redistricting many states so that gerrymandering is avoided to help ensure competitive elections.

Keywords: gerrymandering, redistricting, convexity ratio, district shape, endogon, convex hull.

Optimal Sustainable Constant Effort Fishing Policies in Random Environments

Carlos A. Braumann[1,2], Nuno M. Brites[1]

[1]Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Portugal, [2]Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora, Portugal

We use a general Stochastic Differential Equation model for the growth of a fish population in a randomly varying environment, to which we subtract a harvesting yield term based on a constant or variable fishing effort. We consider a quite general profit structure with linear prices per unit yield and linear costs per unit effort. Previous work on the optimal design of the fishing policy with the purpose of maximizing the expected accumulated profit (discounted by a depreciation rate) over a finite time horizon lead to fishing efforts that vary with the randomly varying population size (sometimes even in a bang-bang way). These policies are not applicable since they need constant evaluation of population size and require very frequent randomly determined changes in the fishing effort. Our approach uses instead a very easily applicable constant effort fishing policy, which leads to sustainability of the population and to a stationary distribution of the population size. We determine the constant fishing effort that optimizes the expected sustainable profit per unit time. Then, for the logistic and the Gompertz models, we use Monte Carlo simulations and check what we lose profitwise by using this policy instead of the optimal inapplicable policy with variable effort; for common situations, our approach is almost as profitable as the first.

Keywords: Fishing Policies, Stochastic Differential Equations, Random Environments, Profit Optimization.

Acknowledgements: *The authors belong to the Centro de Investigação em Matemática e Aplicações (CIMA, ref. UID/MAT/04674/2013), a research centre supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal). Nuno M. Brites holds a PhD grant from FCT (ref. SFRH/BD/85096/2012).*

Taylor's Law via Ratios, for Some Distributions with Infinite Mean

Mark Brown
Dept of Statistics, Columbia University

Taylor's law (TL) originated as an empirical pattern in ecology. In many sets of samples of population density, the variance of each sample was approximately proportional to a power of the mean of that sample. In a family of nonnegative random variables, TL asserts that the population variance is proportional to a power of the population mean. TL, sometimes called fluctuation scaling, holds widely in physics, ecology, finance, demography, epidemiology, and other sciences, and characterizes many classical probability distributions and stochastic processes such as branching processes and birth-and-death processes. We demonstrate analytically for the first time that a version of TL holds for a class of distributions with infinite mean. These distributions and the associated TL differ qualitatively from those of light-tailed distributions. Our results employ and contribute to methodology of Albrecher and Teugels (2006) and Albrecher, Ladoucette and Teugels (2010). This work opens a new domain of investigation for generalizations of TL.

This work is joint with Professors Joel Cohen and Victor de la Pena.

Sharp Bounds for Exponential Approximations of NWUE Distributions

Mark Brown[1], Shuangning Li[2]
[1]*Dept of Statistics, Columbia University, USA,*
[2]*Dept of Statistics and Actuarial Science, University of Hong Kong, Hong Kong*

Let *F* be an NWUE distribution with mean 1 and *G* be the stationary renewal distribution of *F*. We would expect *G* to converge in distribution to the unit exponential distribution as its mean goes to 1. In this paper, we derive sharp bounds for the Kolmogorov distance between *G* and the unit exponential distribution, as well as between *G* and an exponential distribution with the same mean as *G*. We apply the bounds to geometric convolutions and to first passage times.

Keywords: Exponential approximations, reliability theory, NWUE distributions, Kolmogorov distance.

Asymptotic Analysis and Optimization of Insurance Company Performance

Ekaterina Bulinskaya
Dept of Probability Theory, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Russia

It is well known that one can use different criteria (risk measures or objective functions) in order to evaluate an insurance company performance. The most popular one in the last century was the company ruin probability for Cramer-Lundberg model. However in practice it turned out that the negative surplus level not always leads to bankruptcy, since the company may use, e.g. a bank loan to avoid insolvency. So, there were defined and studied “absolute ruin”, “Parisian ruin”, as well as, “omega models”, in the framework of reliability approach. Being a corporation insurance company is interested in dividend payments to its shareholders. Thus, the so-called cost approach arose in the middle of the last century due to pioneering Bruno de Finetti paper. He proposed to maximize the expected discounted dividends paid out until ruin. Modern period in actuarial sciences is characterized by interplay of financial and actuarial methods leading to unification of reliability and cost approaches. For optimization of a company functioning one can use various tools such as investment, bank loans and reinsurance. We are going to investigate asymptotic behavior of several insurance models and apply the results for establishing the optimal investment and reinsurance strategies. The discrete-time models which are more realistic in some situations will be considered along with continuous-time ones.

Keywords: Asymptotic Behavior, Optimization Criteria, Ruin Probability, Bankruptcy, Dividends, Investment, Reinsurance.

Acknowledgment. *The research was partially supported by RFBR grant 17-01-00468.*

On Quantum Information Characterization of Markov and non-Markov Dynamics of Open Quantum Systems

Andrey Bulinski
Dept of Higher Mathematics, Moscow Institute of Physics and Technology, Russia

The theory of open quantum systems continued attracting much attention during the last decade due to the vast area of applications in various branches of physics and also in biology and chemistry. Treatment of the evolution of open system interacting with environment

as Markovian is a useful idealization valid only when memory effects can be neglected. When the initial states of the system and environment are uncorrelated, then reduced dynamics of the system is implemented by completely positive linear maps which obey the semigroup law. Nowadays the studies are mostly devoted to non-Markovian evolution of open systems and are often related to modern quantum technologies of information processing and storing. We can refer to nice recent review papers by Rivas-Huelga-Plenio and by Breuer-Laine-Piilo for characterizations of quantum non-Markovianity and diverse measures of its manifestation in dynamics. Two main approaches to Markovianity property focus on CP (completely positive) -divisibility by an analogy with positive divisibility of classical inhomogeneous Markov processes, and the lack of backward flow of information into the system, respectively. Initially the presence of backflow was formulated in terms of increasing distinguishability of a pair of evolving states. However, the CP-divisibility requirement is stronger, and there have emerged many other nonequivalent measures of the information backflow. A stronger form proposed by Buscemi-Datta being equivalent to CP-divisibility is not easy to verify. The analysis was mainly confined to quantum systems with finite-dimensional underlying Hilbert space. We will be concerned with generalizations of the results on the open quantum system evolutions to comprise the infinite-dimensional case, including the possibility to consider states of more general von Neumann algebras than all the bounded operators. To this aim we benefit from Shirokov's extensions of various entropic measures of quantum correlations.

Keywords: Quantum Information, Open Quantum Systems, Non-Markovian Dynamics, von Neumann Algebras.

Diversification Analysis in Value at Risk Models under Heavy-Tailedness and Dependence

Suzanne Burns
Imperial College Business School, United Kingdom

This paper analyses the effect of diversification across *N* assets, on the Value at Risk of a portfolio whose dependence is captured by a copula. We investigate risks with Power Law marginals - which satisfy

$$P(X > x) \sim \frac{C}{x^{\zeta}}$$

for large x's, where ζ is known as the tail index and X is the potential loss value, including risks with Student-*t* distributions. The copula we show particular interest in is the Student-*t* as, according to a number of works in the literature, it outperforms other copulas when modeling VaR. The

results involving the Gaussian copula will be also included for comparison reasons. The research can be seen as a continuation of [Mo, 2013], which included diversification analysis on the bivariate Student-t copula and Power-Law marginals, amongst its results. We extend the analysis from bivariate data to include *N* assets, and adjust the ratio which measures diversification optimality from

$$\frac{VaR((X_1 + X_2)/2)}{VaR(X_1)} \quad \text{to} \quad \frac{VaR((X_1 + X_2 + ... + X_N)/N)}{VaR(X_1)},$$

where each X_i for i = 1, 2, …, N, is an investment risk (it is natural to refer to the latter quantities as diversification ratios). The results extend previous reports that diversification becomes suboptimal for risks with Power Law marginals for tail indices $\zeta < 1$. We also note how the effect of different input parameters on the ratio changes according to the value of the tail index ζ. The second part of our paper describes the optimizer created to determine numerically the portfolio weights that minimize the VaR for each distribution. The output weights are on par with our results on diversification. Finally, we take the perspective of a portfolio manager that must make assumptions on ζ, and other distribution parameters, and compare the VaR calculated under the assumptions to the VaR obtained knowing the correct distribution.

Empirical Power Study of the Jackson Exponentiality Test

Frederico Caeiro, Ayana Mateus
Faculdade de Ciências e Tecnologia & Centro de Matematica e Aplicações (CMA), Universidade Nova de Lisboa, Portugal

Since many statistical methods depend on the exponential assumption, testing exponentiality has an important role in Statistics. Possible alternative models, which extend the exponential distribution, are the gamma distribution, the Weibull distribution, or the generalized Pareto distribution. Many tests have been proposed in the literature and here we consider the Jackson exponentiality test. In this paper we use Monte Carlo computations to study the Empirical Power of the Jackson test.

Keywords: Exponential distribution; exponentiality test; monte carlo simulation; power of a statistical test.

Reference: O. A. Y. Jackson. An analysis of departures from the exponential distribution, J. Roy. Statist. Soc. B, 29, 540-549, 1967.

Probability Weighted Moments Method for Pareto distribution

Frederico Caeiro, Ayana Mateus
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal and Centro de Matematica e Aplicações (CMA)

The Pareto distribution has been extensively utilized for modelling events in fields such as bibliometrics, demography, insurance, or Finance, among others.

There are several methods to estimate the parameters of the Pareto distributions (see (Arnold [1], Johnson [3], Quandt [4] and references therein). In this work, we consider the probability weighted moments (PWM) method, a generalization of the classic method of moments. In this method we work with the theoretical moments $M_{p,r,s}E\left(X^p\left(F(X)\right)^r\left(1-F(X)\right)^s\right)$ with p, r and s any real numbers, and with their corresponding sample moments. The PWM estimators are obtained by equating $M_{p,r,s}$ with their corresponding sample moments, and then solving those equations in order of the parameters. The PWM estimators for the Parameters of the Pareto distribution were presented in [2]. We now study the performance such estimators for finite sample sizes and compare them with the moment and maximum likelihood methods through a Monte Carlo simulation.

Keywords: Pareto distributions, Probability weighted moments, Monte Carlo simulation.

References:

[1]. B.C. Arnold. Pareto and Generalized Pareto Distributions. In D. Chotikapanich (ed.) Modeling Income Distributions and Lorenz Curves, Springer, New York, 119-145, 2008.

[2]. F. Caeiro and M. I. Gomes. Semi-parametric tail inference through probability weighted moments. Journal of Statistical Planning and Inference, 141, 937-950, 2011.

[3]. N.L. Johnson, S. Kotz and N. Balakrishnan. Continuous Univariate Distributions, Wiley, New York, Vol.1, 2nd Ed., 1994.

[4]. R. E. Quandt. Old and new Methods of Estimation and the Pareto Distribution. Metrika, 10, 55-82, 1966.

[5]. J. Hosking, J. Wallis and E. Wood. Estimation of the generalized extreme value distribution by the method of probability-weighted moments. Technometrics, 27(3),251-261, 1985.

Option Pricing and Model Calibration under Multifactor Stochastic Volatility and Stochastic Interest Rate – an Asymptotic Expansion Approach

Canhanga B.[1], Malyarenko A.[2], Ni Y.[2], Rancic M.[2], Silvestrov S.[2]
[1]Faculty of Sciences, Dept of Mathematics and Computer Sciences, Eduardo Mondlane University, Mozambique,
[2]Division of Applied Mathematics, School of Education, Culture and Communication, a Mälardalen University, Sweden

Among other limitations, the celebrated Black-Scholes option pricing model assumes constant volatility and constant interest rates, which is not supported by empirical studies on for example implied volatility surfaces. Studies by many researchers such as Heston in 1993, Christoffersen in 2009, Fouque in 2012, Chiarella-Ziveyi in 2013, and the authors' previous work removed the constant volatility assumption from the Black-Scholes model by introducing one or two stochastic volatility factors with constant interest rate. In the present paper we follow this line but generalize the model by considering also stochastic interest rate. More specifically, the underlying asset process is governed by a mean-reverting interest rate process in addition to two mean-reverting stochastic volatility processes of fast and slow mean-reverting rates respectively. The focus is to derive an approximating formula for pricing the European option using a double asymptotic expansion method, and present a calibration procedure which is then applied to Swedish option market data.

Keywords: stochastic volatility, asymptotic expansion, stochastic interest rate, European options, calibration.

Online Robust PCA

Hervé Cardot[1], Antoine Godichon-Baggioni[2]
[1]Institut de Mathématiques de Bourgogne, Université de Bourgogne Franche-Comté, France,
[2]INSA Toulouse, Département Génie Mathématique et Modélisation, France

In the current context of data explosion, online techniques that do not require to store all the data in memory are needed to perform principal components analysis (PCA) of streaming data as well as massive data.
Recursive approaches, which are extremely fast and do not require to store all the data in memory, also allow for automatic update when the data are observed sequentially.
For multivariate data, the mean vector and the covariance matrix are classical indicators of central location and multivariate dispersion that can be estimated sequentially. However outlying data may be hard to

detect automatically in large samples multivariate data context and both the mean vector and the covariance matrix can be highly affected by a small proportion of outlying observations.
We introduce in this talk robust indicators of central position and multivariate dispersion based on the geometric median and the median covariation matrix which are relevant objects to perform robust PCA. We explain how such indicators, which can be expressed as the solutions of convex optimization problems, can be efficiently estimated in a recursive and very fast way thanks to averaged stochastic gradient algorithms. The robust principal components can also be simply updated at each new observation.
Numerical experiments on simulated as well as real high dimensional datasets confirm the effectiveness of this online estimation procedure. A comparison to more classical robust PCA techniques confirms the interest of this online robust PCA approach in the presence of outliers.

Keywords: Geometric median, median covariation matrix, stochastic gradient algorithms, updating scheme.

Reference: Cardot, H. and Godichon-Baggioni, A. (2017). Fast estimation of the median covariation matrix with application to online robust principal components analysis. *To appear in TEST*.

Modelling Grade Seniority in Manpower Planning: Markov or semi-Markov?

Philippe Carette[1], Marie-Anne Guerry[2]
[1]Dept of General Economics, Ghent University, Belgium,
[2]Dept Business Technology and Operations, Vrije Universiteit Brussel, Belgium

We consider manpower systems divided into a number of categories (grades), where employees move between the grades over time. If the propensity to change grade is affected by the length of stay in the current grade, the basic Markov model with the categories as state space is not appropriate to describe the system.
One way to deal with this source of heterogeneity is to divide the grades into subcategories that are homogeneous regarding the transition probabilities, so that the Markov model can still be used. The subdivision of the grades is based on a suitable length of stay criterion.
Another approach is to keep the original set of grades and build a semi-Markov model, where the transition probabilities between the grades are allowed to vary with seniority in the grade that is left.
We discuss the pros and cons of the two modelling approaches and compare them in terms of relative quality and internal validity. We hereby fit our models to longitudinal career data on academic staff in a university.

Keywords: Markov model, semi-Markov, manpower planning, seniority.

Multivariate L-moments Statistical Inference, with Hydrological Applications

Fateh Chebana
Centre Eau Terre Environnement,
Institut National de la Recherche Scientifique (INRS), Canada

L-moments, in the classical univariate setting, are widely and commonly used in hydrological applications within a variety of statistical tools. In this talk, we present new developments of statistical tools and methods based on the multivariate version of the L-moments. The motivation behind this direction is that hydrological extreme events, such as floods or droughts, are characterized by several dependent random variables. Hence, the corresponding risk requires to be better evaluated in the multivariate setting and by making the most of all advantages of the L-moments. First, we present a new method to estimate parameters of multiparameter copula. One the important findings of this part is the connection between copulas and L-moments. Second, we developed goodness-of-fit tests specific for multiparameter copulas. Finally, we present multivariate discordancy and homogeneity tests in order to form regions (a set of gauged sites) to make estimations at ungauged sites (without available data). Some asymptotic results, simulation studies as well as case studies are presented, especially in the hydrological context.

Keywords: Multivariate L-moments, goodness-of-fit, parameter estimation, multiparameter copula, flood, homogeneity test.

this work represents several papers in collaboration with B. Brahimi, I. Ben Nasr, A. Necir and T.B.M.J. Ouarda

Using Scan Statistics for the Change Detection in Granger Causality

Jie Chen[1], Thomas Ferguson[1], Paul Jorgensen[2]
[1]University of Massachusetts Boston, USA,
[2]The University of Texas Rio Grande Valley, USA

Scan statistics are defined by the maximum number of events observed in a window. They have been widely used for testing the null hypothesis of a homogeneous distribution against the alternative of clustering in a sub-sequence. If the length and location of clustering are both unknown, variable window scan statistics are recommended. In this talk, we

present variable window scan statistics based on minimum P values statistics (Chen and Glaz 2016) and generalized likelihood scan statistics (Nagarwalla 1996) for the continuous Poisson Process. These scan statistics can be used to monitor data, modeled by stationary time series, to detect the clustering of multiple changes of Granger causality using recursive rolling windows. Simulation studies are presented to evaluate the accuracy of achieving the targeted significance level and to compare the powers of the test statistics mentioned above for a range of alternatives.

Keywords: Granger causality, Likelihood ratio test, Minimum p-value statistic, rolling windows, Simulation studies, Testing homogeneity, Scan statistics, Variable window.

Non-Convex Structured Robust PCA

Stephane Chretien
National Physical Laboratory, United Kingdom

Robust PCA is a technique for data analysis which consists in decomposing the data matrix into the sum *L+S* of a low rank and a sparse matrix. The low rank component is often justified by the same considerations involved in standard PCA analysis which assumes that the information is contained in a small dimensional space. The sparse component represents the outliers. Oftentimes however, some additional structure is available on the sparsity of *S*, like being banded, being columnwise sparse, or being the adjacency matrix of a tree or some other graph. Moreover, Robust PCA is usually addressed using convex optimization approaches such as Semi-Definite Programming. In the present work, we show how these additional structures can easily be incorporated into a computationally efficient non convex optimization scheme with provable convergence to the solution. Our approach is based on the recent paper by Netrapalli, P., Niranjan, U. N., Sanghavi, S., Anandkumar, A., & Jain, P. (2014). Non-convex robust pca. In *Advances in Neural Information Processing Systems* (pp. 1107-1115).We provide in particular a simplified approach of the technical results provided in that paper allowing for a flexible adaptation of the method of proof to different structures of sparsity. Some simulations results will be provided with applications to topology estimation in power grids, time series modelling, and outlier extraction in times series.

Keywords: Conference, CMSIM, ASMDA, Demographics Style.

Risk Factors of Severe Cognitive Impairment in the Czech Republic

Kornélia Cséfalvaiová, Jitka Langhamrová
Dept of Demography, Faculty of Informatics and Statistics, University of Economics, Czech Republic

Expected dramatic increase in the number of people with cognitive impairment will put high demands on health and social care in the Czech Republic. Population aging and the increase of elderly persons aged 65+ evoked a need to address this issue, since age is the major risk factor for dementia and severe cognitive impairment. Conflicting conclusions of European studies confirm the difficulties of quantifying the disease. This article includes the analysis of risk factors of severe cognitive impairment, based on socio-demographic and health variables in the Czech Republic. The method of logistic regression was used for the analysis of risk factors.

Keywords: Population Ageing, Severe Cognitive Impairment, Risk Factors, Czech Republic.

Robust Ranking via Eigenvector and Semidefinite Programming Synchronization

Mihai Cucuringu
Dept of Statistics and the Mathematical Institute, University of Oxford and the Alan Turing Institute, United Kingdom

We consider the classic problem of establishing a statistical ranking of a set of n items given a set of inconsistent and incomplete pairwise comparisons between such items. Instantiations of this problem occur in numerous applications in data analysis (e.g., ranking teams in sports data), computer vision, and machine learning. We formulate the above problem of ranking with incomplete noisy information as an instance of the group synchronization problem over the group SO(2) of planar rotations, whose usefulness has been demonstrated in numerous applications in recent years. Its least squares solution can be approximated by either a spectral or a semidefinite programming (SDP) relaxation, followed by a rounding procedure. We perform extensive numerical simulations on both synthetic and real-world data sets (Premier League soccer games, a Halo 2 game tournament and NCAA College Basketball games) showing that our proposed method compares favorably to other algorithms from the recent literature.
We propose a similar synchronization-based algorithm for the rank-aggregation problem, which integrates in a globally consistent ranking

pairwise comparisons given by different rating systems on the same set of items. We also discuss the problem of semi-supervised ranking when there is available information on the ground truth rank of a subset of players, and propose an algorithm based on SDP which recovers the ranks of the remaining players. Finally, synchronization-based ranking, combined with a spectral technique for the densest subgraph problem, allows one to extract locally-consistent partial rankings, in other words, to identify the rank of a small subset of players whose pairwise comparisons are less noisy than the rest of the data, which other methods are not able to identify.

Keywords: ranking, spectral methods, semidefinite programming, group synchronization.

The Impact of Mortality Projection Models in Case of Flexible Retirement Schemes

Mariarosaria Coppola[1], Maria Russolillo[2], Rosaria Simone[1]
[1]Dept of Political Sciences, Federico II University, Italy,
[2]Dept of Statistics and Economics, University of Salerno, Italy

The trend of mortality is uncertain and this uncertainty causes the so called Longevity Risk. This risk has become one of the key risks that Governments need to manage, not only in the context of welfare policies but also as regards the national Social Security System. The longevity risk from an individual point of view represents the risk that individual mortality rates differ from the expected ones. From an aggregate point of view it represents the risk that unexpected lifestyle changes or medical progress can improve longevity. We can refer to the last one as Trend risk, which has the nature of a systematic risk. Therefore it is not diversifiable and represents a crucial risk component in Social Security Systems management, as well as in pension funds and annuity provider's risk management processes. In this context, many countries have set up Social Security Systems which link retirement age and/or pension benefits to life expectancy, considering a mechanism for indexing the retirement age and/or pension benefits. The issue is a subject of great interest in recent literature; the debate outlines new directions in pension scheme developments and presents experiences with flexible pension schemes from various countries.
In this context, we consider an indexing mechanism based on the expected residual life expectancy to adjust the retirement age and keep a constant Expected Pension Period (EPP). The motivation is to focus on the recent and spread need to create flexible retirement schemes for facing global ageing and the prolonging working lives.

We compare the cost of the Social Security System in case of a retirement age set at a certain age (traditional system) and the cost in the case of the indexed retirement age. In this paper we evaluate the impact on those costs of different selected mortality rates projection models. Empirical evidences are provided.

Keywords: Longevity risk, mortality projections, mortality-indexed life annuities.

"Money Purchase" Pensions: Modeling non Traditional Life Insurance Products

Valeria D'Amato[1], Emilia Di Lorenzo[2], Marilena Sibillo[1], Roberto Tizzano[2]

[1]Dept of Economics and Statistics, University of Salerno, Italy,
[2]Dept of Economic & Statistical Sciences, University of Napoli "Federico II", Italy

In the paper the Authors propose a new personal pension product within the non-traditional profit sharing life insurance contracts, also in light of the indications from the main Authorities involved in the life insurance field, of protecting the members of insured pools from the volatility of long-term returns. In the new contract, the profit participation is structured taking into account both the counterparties' interests and allows the insured to benefit of the profit sharing all along the contract duration, this meaning from the issue time till the insured's death. In its concrete realization, the idea comes true as a sequence of premiums characterized by a level cap, followed by the sequence of benefits characterized by a level floor. The two embedded options are inserted in the basic structure of a pension annuity. Numerical applications are presented, to the aim of investigating the answers of the proposed contract developing the product performance analysis.

Keywords: pension, variable annuity, performance analysis.
JEL Codes: C53, G17, G22, G32

Hitting Times for Claim Number in Car Insurance Setting

Guglielmo D'Amico[1], Fulvio Gismondi[2], Jacques Janssen[3], Raimondo Manca[4], Filippo Petroni[5], Dmitrii Silvestrov[6]

[1]Dip. di Farmacia, Universita "G. d'Annunzio" di Chieti-Pescara, Chieti, Italy, [2]University "Guglielmo Marconi", Italy, [3]Honorary professor at the Solvay Business School Universitè Libre de Bruxelles, Belgium, [4]Dip. di Metodi e Modelli per l'Economia, la Finanza ed il Territorio Università di Roma La Sapienza, [5]Dip. di Scienze Economiche ed Aziendali, Universita di Cagliari, Italy, [6]Department of Mathematics, Stockholm University, Sweden.

In this paper, the phase space of non-homogeneous semi-Markov processes is constructed taking into account the number of claims that an insured will have during her/his driving life. The aim is the calculation, for a driver, of the mean time to report a given number of claims. This problem can be solved constructing the probability distribution function of the first entry time for each state (number of claims) of the model. The age is considered as the non-homogeneous time variable.
As well known, the age in car insurance contracts plays a fundamental relevance in the calculation of behaviour of insured people. In this study, non-homogeneous semi-Markov models will be used for following the time evolution of the claim number.

Keywords: random operators, deterministic barriers, driver reliability, level crossing orders.

Modeling Trading Duration, Volume and Returns by Means of Vector Indexed semi-Markov Chains

Guglielmo D'Amico[1], Filippo Petroni[2]

[1]Dipartimento di Farmacia, Universita "G. d'Annunzio" di Chieti-Pescara, Italy, [2]Dipartimento di Scienze Economiche ed Aziendali, Universita di Cagliari, Italy

The aim of this work is to advance a new stochastic model for describing the dynamic of trading duration, volume and returns. The model is a vectorial extension of Indexed Semi-Markov Chains and is used to investigate the dependence relation existing among the considered financial variables. The methodology is applied to a sample of real high frequency financial intraday data.

Keywords: Indexed Semi-Markov Chains, market microstructure, copula.

Optimal Provision of a Dispatchable Energy Source for Wind Energy Management: Dependence on the Wind Energy Model

Guglielmo D'Amico[1], Filippo Petroni[2], Robert Adam Sobolewski[3]
[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,
[2]Dept of Business and Economics, University of Cagliari, Italy,
[3]Dept of Power Engineering, Photonics and Lighting Technology, Bialystok University of Technology, Poland

Wind energy is assuming even more importance in the production of electricity. The share of production due to wind is continuously increasing in time although there are still relevant problems that affect this industry. The most important limitation for a further development of the wind energy industry concerns the variability of the wind speed phenomenon.
The problem of the wind speed volatility has been approached mainly by energy storage systems; that is by storing a surplus of energy to be used for compensating an eventual future deficit of production. More recently an insurance contract between the wind energy producer (WEP) and a dispatchable energy producer (DEP) has been proposed as a mean to manage the uncertainty of the wind speed.
In this paper we assume that the WEP is also able to produce energy by means of gas and that he has agreed to furnish a given quantity of energy K. An insufficient production of energy determines a cost to be suffered because penalties apply. However an excess of production is lost. Therefore the energy producer should determine the optimal quantity of energy to be produced with gas that added to the uncertain wind energy production maximize his expected profit. The problem is solved under different hypothesis on the wind energy model. First, the wind energy production is modelled by a simple sequence of i.i.d. random variables, then a Markov chain model is used and finally semi-Markov based models of wind energy are applied. The results show the dependence of the optimal policy on the different models of wind energy and therefore highlight the importance of using an appropriate model of wind energy. The application is performed on real data of energy produced by a wind turbine E-48 ENERCON of rated power 800kW.

Keywords: wind energy, optimal provision, semi-Markov.

Dynamic Measurement of Poverty: Modeling and Estimation

Guglielmo D'Amico[1], Philippe Regnault[2]
[1]*Dipartimento di Farmacia, Università "G. d'Annunzio" di Chieti-Pescara, Italy,*
[2]*Laboratoire de Mathématiques de Reims, Université de Reims Champagne-Ardenne, France*

This study presents a model of income evolution from which dynamic versions of commonly used static poverty measures are derived. The dynamic indexes are calculated both for fi nite and infi nite size economic systems. Estimation based on micro-data and macro-data is also discussed under di fferent sampling schemes and it is proved that estimators are strongly consistent. Simulations are performed so as to illustrate the computability and interpretability of our indexes and their estimates.

Keywords: Markov process; population dynamic; parametric estimation; poverty index.

Financial Risk Distribution in European Union

Guglielmo D'Amico, Stefania Scocchera, Loriano Storchi
Dipartimento di Farmacia, Università "G. d'Annunzio" di Chieti-Pescara, Italy

The aim of this work is to assess the distribution of financial risk within 26 European Countries and to determine the inequality of it using data from January 1998 to November 2016. The data are: sovereign credit ratings assigned by Moody's, Standard & Poor's and Fitch along with interest rates on a monthly time scale collected by European Central Bank and Federal Reserve Bank website. From these data we recover the credit spread distributions depending on the rating classes they refer to. The model consists of a discrete-time homogeneous Markov chain for the credit rating dynamic and a reward process for the credit spreads. The inequality of financial risk distribution is estimated by means of Dynamic Theil Entropy which allows to forecast future inequality for next three years. The methodology is applied to several subsamples of Countries in order to better understand the variation of this inequality in case of some Countries leave out European Union.

Keywords: Markov Chains, sovereign credit ratings, credit spreads, Dynamic Theil Entropy.

Optimal Portfolio Strategies and Derivative Products under Insider Information

Bernardo D'Auria[1], Carlos Escudero Liebana[2]
[1]*Statistics Dept, Madrid University Carlos III, Avda. Universidad, Spain,*
[2]*Mathematics Dept, Madrid University Autonoma, Spain*

The optimal portfolio for a small investor is a classical problem that has attracted a lot of interest among (financial) mathematics researchers and practitioners in finance. In this talk we analyze the case in which the investor has access to insider information and we show how to quantify the value of the additional information s/he owns. In particular we focus on the case in which the additional information is not precise, for example assuming that s/he does not know exactly the price of a given asset at a final time T, but s/he knows that it is contained in a, possible discrete, measurable set.

Keywords: Portfolio Optimization, Insider Information, Enlargement of Filtrations.

An Intervention Analysis regarding the Impact of the Introduction of Budget Airline Routes to Maltese Tourism

Maristelle Darmanin, David Suda
University of Malta, Malta

Intervention analysis is an important method for analysing temporary or long-lasting effects of sudden events on time series data. We use monthly data of the National Statistics Office's Tourstat survey covering the years 2003 up to 2012. This contains a number of time series regarding tourist demographics, the type of tourism, the type of accomodation sought, total tourist nights and total expenditure. We apply intervention analysis to determine the impact of the introduction of budget airline routes to these Maltese tourism related time series. We consider two main interventions. The first is the introduction of Italy and UK bound routes in October 2006, Italy and UK being two of Malta's major tourism markets. The second is the introduction of a considerable number of routes in March 2010, in particular the Marseille route. In addition to the standard types of intervention, the step and the pulse intervention, we also introduce a periodic pulse intervention which allows us to cater for any seasonality in the intervention effect, with the corresponding transfer function possibilities. We conclude with a discussion of the adequacy of the fitted models by comparing their dynamics with moving average plots of the original series.

Keywords: Intervention analysis, SARIMA models, tourism statistics

Statistical Inference in a Model of Imperfect Maintenance with Geometric Reduction of Intensity

J.-Y. Dauxois[1], S. Gasmi[2], O. Gaudoin[3]
[1]University of Toulouse-INSA and IMT, Département GMM, France, [2]Université de Tunis, École Nationale Supérieure d'Ingénieurs de Tunis (ENSIT), TUNISIA, [3]Laboratoire Jean Kuntzmann, Tour IRMA, FRANCE

The aim of this paper is to introduce and study a new model of Imperfect Maintenance in Reliability. A model of geometric reduction of intensity is assumed on the inter-arrival times of failures on a system subject to recurrent failures. Based on observation of several systems, we introduce estimators of the parameters (euclidean and functional) of this semiparametric model and we prove their asymptotical normality. Then a simulation study is carried out to learn the behavior of these estimators on samples of small or moderate size. We end this work with an application on a real dataset.

Keywords: Imperfect repair, Failure intensity, Large sample behavior, Reliability, Semiparametric inference.

Improved Bounds for the Probability of Causation

A. Philip Dawid, Monica Musio
University of Cagliari, Dept of Mathematics, Italy

In many applications, such as disputes at Law, interest lies in whether a specific exposure can be regarded as having caused an observed effect. But there are difficulties in addressing such a "Causes of Effects" query using incidence rates obtained from observational or experimental data, which are more suited to addressing general scientific queries about the "Effects of Causes". When the object of concern is a specific individual, it is not always clear how one could usefully employ scientific data to inform inferences about individual events. Indeed, given even the best possible empirical evidence about the probabilistic dependence of the outcome on the exposure, we can typically only provide interval bounds for the "probability of causation" for the case of a specific individual who has developed the outcome after being exposed.
In this work we show how these bounds can be refined if we have further information about internal mechanisms and processes, in the form of additional variables measured in the data. In particular, we show how this can be done using information on covariates, confounders and complete or partial mediators, separately or in combination.

Keywords: Probability of Causation, Causes of Effects, Covariate, Confounder, Mediator.

Multivariate L-moments defined through Transport

Alexis Decurninge
Huawei Technologies Co. Ltd., Mathematical and Algorithm Sciences Lab, France Research Center, France

L-moments are used as alternative to moments for the description of a univariate distribution with the only assumption of finite expectation. Univariate L-moments are expressed as projections of the quantile function onto an orthogonal basis of polynomials in $L_2([0;1],R)$. We present multivariate versions of L-moments expressed as collections of orthogonal projections of a multivariate quantile function on a basis of multivariate polynomials in $L_2([0;1]^d,R)$. Similarly to the univariate case, such multivariate L-moments exist as soon as the expectation of the underlying multivariate distribution is finite and completely characterize this distribution.
Contrary to the univariate case, there is no consensus on the way to define multivariate quantile. We propose to consider multivariate quantile functions defined as transport from the uniform distribution on $[0;1]^d$ onto the distribution of interest, each particular transport leading to a different definition of multivariate L-moments. We will in particular present the case of the transport of Rosenblatt leading to the L-comoments proposed by Serfling and Xiao (2007) and a particular optimal transport defined as the gradient of a convex function.
We will present different ways to envisage the estimation of multivariate L-moments and the asymptotic properties in the case of plug-in estimators associated to the two different transports studying in particular the conditions of their consistency.

Keywords: Multivariate distribution, Quantile, Transport, Dependency structure.

Computing the Mutual Constrained Independence Model

Thomas Delacroix[1], Philippe Lenca[1], Stéphane Lallich[2]
[1]Institut Mines-Telecom, Telecom Bretagne, UMR CNRS 6285 Lab-STICC, France, [2]Université de Lyon, Laboratoire ERIC, France

Developed for applications in itemset mining, the notion of Mutual Constrained Independence is a natural generalization of the notion of mutual independence. If the mutual independence model on a finite number of events can be seen as the least binding model for the

probabilities of any finite intersection of these events, given the probabilities of each of these events, then the Mutual Constrained Independence Model on a finite number of events can be seen as the least binding model for the probabilities of any finite intersection of these events, given the probabilities of any number of such intersections of events.
In this article, we present a first detailed and effective means of computing the Mutual Constrained Independence Model. We show the efficiency of our algorithm and the adequacy of the model by applying it to various examples. A test for the Mutual Constrained Independence Hypothesis is also presented.

Keywords: Independence model, Mutual constrained independence, Itemset mining.

Multivariate European Option Pricing in a Markov-Modulated Lévy Framework

Griselda Deelstra, Matthieu Simon
Service Sciences Actuarielles, Département de Mathématique, Université libre de Bruxelles, Belgium

In this talk, we focus on the pricing of some multivariate European options, namely Exchange options and Quanto options, when the risky assets involved are modelled by Markov-Modulated Lévy Processes (MMLPs). Pricing formulae are based upon the characteristic exponents by using the well-known FFT methodology. We study these pricing issues both under a risk neutral martingale measure and the historical measure. The dependence between the asset's components is incorporated in the joint characteristic function of the MMLPs. As an example, we concentrate upon a regime-switching version of the model of Ballotta et al. (2016) in which the dependence structure is introduced in a flexible way.
Several numerical examples are provided to illustrate our results.

Clustering Variables with Nonlinear Relationships: An Approach based on Polynomial Transformation and a Dynamic Mixed Criteria

Christian Derquenne
Electricité de France, Research and Development, France

The research structures in the data this essential aid to understand the phenomena to be analyzed before any further treatment. Unsupervised

learning and visualization techniques are the main tools to facilitate these research facilities. We have proposed a set of methods for clustering numeric variables in 2016. These are based on a mixed approach: linear correlation test between the variables (initial variables and/or first principal components) and one-dimensionality test (Saporta, 1999) of the resulting groups to dynamically build a typology by controlling the number of classes and quality. It allows primarily to "discover" an "optimal" number of clusters without fixing it a priori. We propose an extension of this approach for nonlinear relationships between variables. But the discovery of clusters is more complex than linear relationships. Indeed, linear correlation test is no longer valid, then we use polynomial model to obtain the relationship level between two variables (initial variables and/or first principal components). In addition, the one-dimensionality test is revisited to adapt to the nonlinearity between the variables. On the other hand, we propose too an extension in presence of outliers and/or missing data in datasets. These two problems are resolved in means of robust tests for the outliers and NIPALS algorithm for the missing data.
Then, as part of energy management, we built time series typologies in the areas of market prices. The characterization of each group of curves obtained allowed to identify and understand the behavior of the joint evolution of the phenomena studied and to detect differences in behavior between clusters.
Lastly, we conclude on future research in frame of high dimension for variables and massive data for individuals. Indeed, the classical statistical tests are no longer valid.

Keywords: Clustering variables, correlation, unidimensionality, nonlinear, outlier, missing data, unsupervised learning.

Distribution of Specific Costs of Agricultural Production in the European Union: an Approach based on the Quantile Regression Method

Dominique Desbois
UMR Economie Publique, INTRA-AgroParistech,
Université Paris-Saclay, France

Introducing the estimate by product of the agricultural production costs according to the quantile regression approach for member countries of the European Union, this paper is structured as follows. After recalling the conceptual framework of the estimation of agricultural production costs, the first part, presents this quantile regression approach, in accordance with the characteristics of the distribution of specific charges for agricultural production, especially its asymmetry. The second part

documents the data collection used by this estimation procedure and distributional characteristics of specific costs for specific production of twelve Member States of the European Union. According to a comparative analysis between the member states, the third part presents the econometric results of products for wheat, dairy milk and pork using factor analysis and hierarchic clustering based on estimation intervals. The last section discusses the relevance of the results obtained.

Keywords: input-output model, agricultural production cost, micro-economics, quantile regression, factor analysis and hierarchic clustering of interval estimates.

References:

[1]. Billard L., Diday E. (2006) *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, 321 p.

[2]. Cazes P., Chouakria A., Diday E., Schektman Y. (1997) Extensions de l'analyse en composantes principales à des données de type intervalle. *Revue de Statistique Appliquée*, n°24, pp. 5-24.

[3]. Chavent M. (2000) Citerion-Based Divisive Clustering for Symbolic Objects, *Analysis of Symbolic Data*, eds. H.H. Bock, E. Diday, Springer, 2000.

[4]. Desbois D., Butault J.-P., Surry Y. (2013) Estimation des coûts de production en phytosanitaires pour les grandes cultures. Une approche par la régression quantile, *Economie Rurale*, n° 333. pp.27-49.

Hazard Rate Estimator for Right Censired Data under Association

Samra Dhiabi[1], Ourida Sadki[2]

[1]Dept of Mathématics Univ. Mohamed Khider, Algeria,
[2]Laboratory RECITS, Faculty of Mathematics, Univ. Sci. and Tech. Houari Boumédiène, Algeria

In this paper we study a smooth estimator of the conditional hazard rate function in the censorship model, when the date exhibit some dependence structure. We show, under some regularity conditions, that the kernel estimator of the conditional hazard rate function is consistent and suitably normalized, it is asymptotically normally distributed.
Somme simulations are drawn to illustrate the main results.

Keywords: Association, Censored data, Conditional hazard rate, Kaplan Meier estimator.

Thinking by Classes and their Symbolic Description in Data Science

E. Diday
CEREMADE Paris-Dauphine University, France

A Data Scientist is someone able to manage and extract new knowledge from any kind of standard, complex or big data. Classes play an important role in Data Science as by clustering they can provide a concise and structured overview on the data and by supervised learning machine they can provide useful decision rules. A third way, is to consider classes as objects by themselves to be described in an explanatory way and then analyzed. Such classes often represent the real units of interest. In order to take variability between the members of each class into account, classes are described by intervals, distributions, set of categories or numbers sometimes weighted and the like. In that way, we obtain new kinds of data, called "symbolic" as they cannot be reduced to ordered numbers without losing much information. Symbolic Data Analysis (SDA) gives answers to big and complex data challenges as big data can be reduced and summarized by explanatory class descriptions and as complex data with multiple unstructured data tables and unpaired variables can be transformed into a structured data table with paired symbolic valued variables. In this talk we focus on a new way to build simultaneously classes and their symbolic description based on mixture decomposition by using Dynamic Clustering (Diday) and Cluster wise of Saporta kind which provide unbiased classes at the contrary of standard mixture decomposition of Dempster kind.

Keywords: Data Science, Data Mining, classification, learning, Symbolic Data Analysis, dynamic clustering, cluster wise, mixture decomposition.

References:

[1]. Dempster A. P., Laird N. M., and Rubin D. B. (1977) Maximum likelihood from incomplete data wia the EM algorithm. Journal of the Royal Statistical Society B, 39(1):1–22, 1977.

[2]. Diday E., Simon J.C. (1980) "Clustering Analysis" -, Chapter in "Communication and Cybernetics Digital Pattern Recognition", K.S. FU edit. Springer Verlag.

[3]. Diday E. (2016) Thinking by classes in Data Science: the symbolic data analysis paradigm. WIREs Comput Stat 2016, 8:172–205. doi: 10.1002/wics.1384

[4]. Preda, C. and Saporta, G. (2005b) “Clusterwise PLS regression on a stochastic process”. Computational Statistics and Data Analysis, 49, pp.99–108.

[5]. Saporta, G. (2008) “Models for Understanding versus Models for Prediction”. Proceedings COMPSTAT'08, Brito, P. (ed.), Springer, pp.315-322.

Optimal Control of a Pest Population through Geometric Catastrophes

Theodosis D. Dimitrakos[1], Epaminondas G. Kyriakidis[2]
[1]*Dept of Mathematics, University of the Aegean, Greece,*
[2]*Dept of Statistics, Athens University of Economics and Business, Greece*

We study the problem of controlling the stochastic growth of a bounded pest population by the introduction of geometric catastrophes. The damage done by the pests is represented by a cost. Another cost is also incurred when the controlling action of introducing geometric catastrophes to the population is taken. It is assumed that the catastrophe rate is constant. We aim to find a stationary policy which minimizes the long-run expected average cost per unit time. A semi-Markov decision formulation of the problem is given. It seems intuitively reasonable that the optimal policy is of control-limit type, i.e. it introduces geometric catastrophes if and only if the pest population is greater than or equal to a critical size. Although a rigorous proof of this assertion is difficult, a computational treatment of the problem is possible. Various Markov decision algorithms are implemented for the computation of the optimal policy. From a great number of numerical examples that we have tested, there is strong evidence that the optimal policy is of control-limit type.

Keywords: Pest control, Geometric catastrophes, Semi-Markov decision process, Control-limit policy, Markov decision algorithms.

Entropic Analysis of Mixture Binomial Distributions applied to Online Ratings

Yiannis Dimotikalis
Dept of Accounting & Finance, T.E.I. of Crete, Greece

In rating data, entropy is a measure of rater's agreement and possibly a criterion of model selection. Using the Shannon entropy, we compare Mixture Binomial and CUB models for their appropriateness to a thousand of real online rating data sets from Amazon, Google Play and TripAdvisor websites. This approach confirmed in real rating data provides the characteristics of different models and explain the behavior of people in rating. Our analysis shows that the fitting performance of the models to real data sets depends on entropy of the data.

Keywords: Shannon Entropy, Mixture Binomial, CUB model, Online Rating, Amazon, Google Play, TripAdvisor.

Asymptotics for a Conditional Quantile Estimator under Censoring and Association

Wafaa Djelladj[1], Abdelkader Tatachak[2]
[1]*Laboratory of MSTD, Dept of Probability and Statistics, USTHB, Algeria,*
[2]*Laboratory of MSTD, Faculty of Mathematics, USTHB, Algeria*

In survival analysis, it is common to deal with sequences of observations that are derived from stationary processes satisfying the association dependence in the sense of Esary et al. (1967). And, due to random right censoring effect, the lifetimes are not completely observed. The main goal of our study in the present work is to assess strong uniform consistency rates and asymptotic normality for a conditional quantile function estimator under association dependency and right censored model. The accuracy of the studied estimates is checked by a simulation study.

Keywords: Association, Censoring, Conditional quantile estimator, Strong uniform consistency.

Development and Application of Multifractal Analysis for EEG Studies in a State of Meditation and Background

Dmitrieva L.A., Zorina D. A, Kuperin Yu. A., Smetanin N. M.
Saint Petersburg State University, Russia

The paper analyzes the EEG time series in the state of background and meditation by methods of multifractal analysis. With the cognitive point of view of particular interest is the impact of meditation on the brain functioning. Total available of EEG recordings were 45 subjects in a state of background and meditation. All subjects were divided into 2 groups: experienced meditators (more than 3000 hours of meditation practice) and inexperienced meditators (less than 300 hours of meditation practice). The goal was to find quantitative differences between groups of experienced and inexperienced subjects. To achieve this goal it was resolved a number of intermediate tasks: data collection and pre-processing; definition of multifractal characteristics and their statistical processing. In the study, we were put features that have not previously been used in the analysis of the EEG time series. Namely, it was decided to calculate 2 quantitative characteristics: the distance from the center of the multifractal spectrum to its left end, i.e the width of the left tail of the spectrum, and the distance from the center of the multifractal spectrum to its right-hand end of the spectrum, that is, the width of the right tail. These characteristics have been calculated for all

multi-channel EEG from an existing database. Statistical processing of the results was performed. It was found that for experienced meditators in meditation state and in background mentioned characteristics significantly different. It was also found that for inexperienced meditators in meditation and background characteristics mentioned statistically indistinguishable

Keywords: multifractal analysis, electroencephalograms, states of meditation and background, statistical analysis.

Modeling of EEG Signals by using Artificial Neural Networks with Chaotic Neurons

Dmitrieva L. A., Kuperin Yu. A., Mokin P. V.
Saint Petersburg State University, Russia

The aim of this paper is to use artificial neural networks in order to generate time series which reproduce the properties of the real electroencephalograms. The focus was made on manifestations of nonlinearity and deterministic chaos. The study deals with a question about the place of modeling in the study of complex systems and processes. Particular attention is paid to the definition of indicators of studied time series. The goal was to make meaningful judgments about the intrinsic properties of the generated EEG signals in comparison with the real signals. We presented and evaluated the simulation results with the help of artificial neural networks. The conclusion is that under certain conditions the neural networks with chaotic neurons may reproduce properties of real EEG signals. But at the same time the similarity between generated and real signal is not so close that there is no way to distinguish one from another by using a sufficiently informative methods including visually representing information. In other words, the concept of cybernetic black box is limited in practice by the complexity of the problem.

Keywords: artificial neural networks, chaotic neurons, electroencephalograms, nonlinear time series, deterministic chaos.

Brownian Motion Exit Densities for General One-Sided Boundaries

Doncho Donchev
Faculty of Mathematics and Informatics,
University of Sofia "St. Kliment Ohridski", Bulgaria

In our recent paper, we characterized the exit density of a Brownian motion for one-sided smooth boundaries in terms of a suitable solution of some parabolic second-order PDE. It turns out that this equation can be reduced to a first-order PDE. It is shown that the last equation admits closed solutions only for three classes of boundaries- parabolic boundaries, square-root boundaries and rational functions. Our approach is substantiated by an example, where we find the exit density for a boundary not studied so far.

A Joint Mixed Model for Longitudinal Data Involving Time-Varying Covariates

Reza Drikvandi
Statistics Section, Department of Mathematics, Imperial College London, UK

Longitudinal and follow-up studies often produce data involving both time-invariant and time-varying covariates. Conventional longitudinal models only model the response variable and ignore the (stochastic) covariate process by treating all covariates as fixed variables, while in addition to the response variable, a time-varying covariate also changes over time and its evolution over time, which provides important information, needs to be modelled as well. Moreover, ignoring the covariate process for time-varying covariates implies that the response variable measured at each follow-up time depends on the time-varying covariates measured at that time point only, but the response variable could also depend on the previous measurements (history) of time-varying covariates. We propose a joint mixed model for the response variable and the time-varying covariates which not only takes into account the covariate process for time-varying covariates but also allows the response variable to depend on the history of time-varying covariates. In our joint modelling approach, the association between the response and the time-varying covariates is taken into account through correlated random effects. We use P-spline functions of time to capture the evolutions of the response and the time-varying covariates over time. Another advantage of our joint model is that it is also applicable to situations where the response and the time-varying covariates for each subject at each follow-up are not measured at the same time. Our joint

modelling approach can help handle missing time-varying covariates and responses simultaneously. The proposed joint mixed model is investigated theoretically and practically, and motivated by data from an AIDS cohort study in which HIV+ subjects have CD4 cell count and viral load measured at repeated visits before and after receiving treatment.

Keywords: Joint model, Longitudinal data, P-spline, Random effects, Time-varying covariate.

Search and Recall: Statistical Learning Theory

Romney B Duffey
DSM Associates Inc., USA

We develop the basis for a new theory to explain how humans learn from experience by correcting mistakes and errors. It is known from millions of data (on accidents, events, crashes, and human trials) that learning curves for individuals and for entire systems are exponential in shape. We need to explain and understand why they are that mathematical shape, and what physically happens inside the brains of individuals as they learn and recall. We propose a new simple statistical theory based on applying classic random search and location theory to human brains.
Externally, it is observed that individual humans follow learning curves for physical tasks, solving problems and improving performance, reducing mistakes, errors and faults by repetition and response. This same learning process and curve also appears at the organizational level. Uncountable internal and unobserved human interactions within the organization or society appear as the external observed events and outcomes and culture. Similarly, when a human searches internally (in what is called memory), there are neural avalanches as the mind seeks these pre-existing or prior experience patterns to permit recognition and recall. This is how we recognize objects and create new ideas and information based on searching, locating, processing and recalling our prior learned experience.
We adopt and adapt Koopman's classic random statistical search theory, originally applied to detection of submarine targets in warfare, to neural search and individual recall and recognition processes. We demonstrate the form of learning curves at the neural level is consistent with the observed behavior and learning curves for both individuals and entire technological systems. This theory therefore unifies modern learning, behavioral and neurological concepts with outcomes and events observed in modern technological systems.

Keywords: Learning curves; theory; search and recall, neural processes.

Stratified Logrank Test under Missing Data

Jean-François Dupuy[1], Rim Ben Elouefi[1,2]
[1]Mathematics Research Institute and Institut National des Sciences Appliquées, France, [2]Computational Mathematics Lab, Faculty of sciences, Monastir University, Tunisia

The stratified logrank test can be used to compare survival distributions of several groups of patients, while adjusting for the effect of some discrete variable that may be predictive of the survival outcome. In practice, it can happen that this discrete variable is missing for some patients. An inverse-probability-weighted version of the stratified logrank statistic is introduced to tackle this issue. Its asymptotic distribution is derived under the null hypothesis of equality of the survival distributions. A simulation study is conducted to evaluate the proposed test statistic in finite samples, in particular against an accelerated failure time model alternative. An analysis of a medical dataset illustrates the methodology.

Keywords: Inverse-probability-weighting, AFT alternative.

Modelling a Dynamic Size Biased Sampling

P. Economou[1], G. Tzavelas[2], A. Batsidis[3]
[1]Dept of Civil Engineering, University of Patras, Greece,
[2]Dept of Statistics and Insurance Science, University of Piraeus, Greece,
[3]Dept of Mathematics, University of Ioannina, Greece

Size biased sampling refers to a type of nonrandom sampling method in which the probability of a population unit to be included in a sample is proportional to some nonnegative weight function w(x) of its size x. A characteristic example of such a situation is the measurements derived from patient admissions to a hospital. In such a case is natural to assume that individuals with severe symptoms (associated for example with high values of virus loads) are more likely to visit a hospital for diagnosis or treatment than a person with less severe symptoms. This implies that the medical personnel observe a size biased sample (in this case a length biased sample) in which the probability of an individual's inclusion in a sample is proportional to its virus load X. On the other hand, one could expect that the biasness in the observed sample from a finite population reduces as the sample size increases, for example during a flu outbreak. This implies that we do not have a full control on the sampling mechanism and more specifically that we cannot assume a stable, supervised sampling mechanism.
In the present work we propose a model to describe situations in which the sampling mechanism is determined and controlled by the phenomenon itself.

Keywords: Biasness, Finite Population, Weighted Distributions.

SNIPE for Memory-Limited PCA with Incomplete Data: From Failure to Success

Armin Eftekhari, Laura Balzano, Michael B. Wakin, Dehui Yang
Alan Turing Institute, United Kingdom

Consider the problem of identifying an unknown subspace S from data with erasures and with limited memory available. To estimate S, suppose we group the measurements into blocks and iteratively update our estimate of S with each new block.
In the first part of this talk, we will discuss why estimating S by computing the "running average" of span of these blocks fails in general. Based on the lessons learned, we then propose SNIPE for memory-limited PCA with incomplete data, useful also for streaming data applications. SNIPE provably converges (linearly) to the true subspace, in the absence of noise and given sufficient measurements, and shows excellent performance in simulations.

Price Sensitivities for Stochastic Volatility Models

Youssef El-Khatib[1], Abdulnasser Hatemi-J[2]
[1]Dept of Mathematical Sciences, UAE University, United Arab Emirates,
[2]Dept of Economics and Finance, UAE University, United Arab Emirates

We deal with the calculation of price sensitivities for stochastic volatility models. We consider general forms for the dynamics of the underlying asset price and its volatility. We make use of Malliavin calculus to compute the price sensitivities.
Obtained results are applied to several recent stochastic volatility models as well as existing ones that are commonly used by practitioners. Each price sensitivity is a source of financial risk. The suggested formulas are expected to improve on hedging of the underlying risk.

Keywords: Asset Pricing, Malliavin Calculus, Price sensitivity, Stochastic volatility, Risk management, European options.

A Discrete Piecewise Multi-state Survival Model: Application to Breast Cancer

Juan Eloy Ruiz Castro[1], Mariangela Zenga[2]
[1]Department of Statistics and Operational Research, Faculty of Science, Campus Fuentenueva s/n University of Granada, Spain, [2]Department of Statistics and Quantitative methods, University of Milano-Bicocca, Italy

Multi-state models are considered in the survival field for modeling illnesses which evolve through several stages over time. Multi-state models can be developed by applying several techniques, such as non-parametric, semi-parametric and stochastic processes, Markov processes in particular. When the development of an illness is being analyzed, its progression is tracked in a periodic form. Medical reviews take place at discrete times and a panel data analysis can be formed. In this paper, a discrete non-homogeneous piecewise Markov process is constructed for modeling and analyzing a multi-state illness with a general number of states. The model is built, the likelihood function in different cases and relevant measures, such as survival function, transition probabilities, mean total times in a group of states and the conditional probability of state change are determined. Covariates depending on time are introduced in the modeling, the results are obtained in a matrix algebraic form and algorithms are shown. The model is applied for analyzing the behavior of breast cancer. A study of the relapse and survival times for 300 breast cancer patients who have undergone mastectomy is developed. The results of this paper were implemented computationally with MATLAB and R.

Keywords: Survival, breast cancer, piecewise Markov model, multi-state model.

PageRank re-Calculation Methods based on Specific Types of Changes in a Graph

Christopher Engström, Sergei Silvestrov
Division of Applied Mathematics, Education, Culture and Communication (UKK), Mälardalen University, Sweden

PageRank is a method used mainly to rank home pages on the Internet by considering a kind of random walk on the Web graph constructed in such a way that home pages and links between different home pages are the respective vertices and edges of the graph.
The Web graph is both very large and constantly changing, this calls for efficient methods not only to calculate PageRank but also to re-calculate PageRank as the graph changes. In this paper we will focus on small

localized changes in the graph such as a small change in a single strongly connected component in the graph.
By considering a non-normalized definition of PageRank rather then normalized PageRank as defined by S. Brin and L. Page we will show how the difference in rank can be calculated both theoretically as well as numerical experiments. In particular we will consider small changes such as the addition or deletion of some edges in a strongly connected component of the graph.

Keywords: PageRank, random walk, graph.

Portfolio Strategies and Filtering within Regime-Switching Models

Christina Erlwein-Sayer[1], Stefanie Grimm[2], Peter Ruckdeschel[3], Joern Sass[4], Tilman Sayer[5]
[1]OptiRisk Systems, United Kingdom, [2]Fraunhofer Institute for Industrial Mathematics, Germany, [3]University of Oldenburg, Germany, [4]University of Kaiserslautern, Germany, [5]Advanced Logic Analytics, United Kingdom

Our asset allocation model is set in a regime-switching market allowing parameters of asset returns to adapt to market changes. Assets are modeled through a multivariate hidden Markov model (HMM) in discrete time with switching drift and volatility goverened by filtered market states. We consider different parametrisations of the involved asset covariances, allowing them to be either independent of the regime or led by filtered states. We utilize a filter-based EM-algorithm, which was pioneered by Elliott (1994) to find adaptive parameter estimates of the assets‘ drift and volatiliy in this multivariate HMM. Our portfolio strategies are based upon the estimated asset distributions. A simulation study as well as a study on actual data show that our strategies outperform naive strategies and strategies with no regime-switching. Furthermore, we enhance long-short trading strategies by making proportion decisions depending on the current estimated regime.

Keywords: Regime-switching model, Filtering, Asset allocation, Trading strategies.

Statfda, an Easy to Use Tool for Functional Data Analysis without Expert Knowledge

M. Escabias, Ana M. Aguilera,
M. C. Aguilera-Morillo, M. J. Valderrama
University of Granada, Dept of Statistics, Faculty of Pharmacy, Spain

The objective of this Special Session is to present Statfda. Statfda is a web based application to use Functional Data Analysis methods in an easy to use mode without a deep knoledgement of the underlying methodology. The generic name of functional data analysis, popularized by Ramsay and Silverman in 1997, encompas a set of statistical methods where individual observations are curves of differente origin, as curves of temperature, spectrometry or kinetics, instead of single observations of a variable. The majority of statistical methods have been developed in the field of functional data analysis as principal component analysis, regression methods, and so on. The problem is that these methods are not included in the most general statistical software as SPSS, SAS, STATA, etc or need to know some technical details, that make difficult for applied researchers to use them. Most of FDA methods are available in R, S-Plus or Matlab thank to scientific comunity. Unfortunately applied researchers of different fields like medicine, chemistry, sport sciences... are lack of expert knowledge of this kind of software so that is difficult for them to take advantage of these methods. With Statfda we have tried to help applied researches to use some useful functional data methodologies as basic exploratory analysis for functional data, functional principal component analysis, functional linear regression and functional logit regression, without too much knowledge of the statistical methodology and R software.

The application is web based developed so that the user do not need software instalation to use it. Statfda is moreover based on R language that is the most used one in functional data analysis.

This session is aimed to all kind of researchers, those who use statistical methods from applied point of view and expert statisticians. After a brief introduction to functional data analysis, we will teach to use the application, the format of the input information and the different output we can get. The session is programmed as a workshop where all the attendees will be able to use the application and learn to manage with it. We will prepare some different classical data sets of the functional data analysis field and will drive attendees throw the steps to get each one of the programmed functional methods. To this aim it will be necessary that participants have their own laptops.

Random Network Evolution Models

István Fazekas, Csaba Noszály,
Attila Perecsényi, Bettina Porvázsnyik
University of Debrecen, Faculty of Informatics, Hungary

To describe the evolution of networks the preferential attachment model was proposed by Barabási and Albert [1]. It is known that the preferential attachment model results in a scale-free random graph. A random graph is called scale-free if it has a power law degree distribution. There are several versions of the preferential attachment model.
In this paper we study network evolution procedures which combine the preferential attachment and the uniform choice. In our models certain parts of the network are characterized by weights. During the evolution both the size of the network and the weights are increased. We prove that the weight distributions are scale-free. We also study the degree distributions.

Keywords: network, preferential attachment, scale free

References:

[1]. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science 286, no. 5439, 509–512, 1999.
[2]. I. Fazekas, Cs. Noszály, A. Perecsényi. Weights of cliques in a random graph model based on three-interactions. Lith. Math. J. 55, no. 2, 207–221, 2015.
[3]. I. Fazekas and B. Porvázsnyik. Scale-free property for degrees and weights in a preferential attachment random graph model. J. Probab. Stat. Art. ID 707960, 12 pp. 2013.
[4]. I. Fazekas and B. Porvázsnyik. Limit theorems for the weights and the degrees in an N-interactions random graph model. Open Mathematics 14 (1), 414-424, 2016.

Decomposition of Marital Status Differences in Life Expectancy by Age in the Czech Republic

Tomas Fiala, Jitka Langhamrova
University of Economics, Czech Republic

Differences in life expectancies by marital status is a well-known phenomenon. To understand better the nature of these differences it is appropriate to use the decomposition method which can detect contributions of individual age groups in adult age to the total difference.
The aim of this paper is to quantify the age-specific contributions of individual age groups to the total differences in life expectancy at births by marital status in the Czech Republic since 1990. The analysis is based on annual Czech Statistical Office data. Each five-year time period is analyzed separately.
Highest contributions to the total differences in life expectancy at births were observed for males mainly in the age groups 50–59 and 60–69

years, for females usually in the age groups 60–69 and 70–79 years. In first decade studied for males under 60 single the highest contribution show single males while at higher ages divorced and widowed. In the last decade, single males have highest contribution in all age groups. For females, the development is more regular and stable. The highest contribution is observed for single females, the lowest for widowed.

Keywords: mortality, family status, decomposition by age, the Czech Republic.

Prevalence of Pediatric High Blood Pressure: a Preliminary Estimate

M. Filomena Teodoro[1,2], Carla Simão[3,4]
[1]CINAV, Center of Naval Research, Naval Academy, Portugal, [2]CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, Lisbon University, Portugal, [3]Faculdade Medicina, Lisbon University, Portugal, [4]Departamento de Pediatria do Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Portugal

A study about pediatric hypertension is introduced. A questionnaire is designed to be answered by the caregivers of Portuguese children and teenagers.
The collected data is statistically analyzed, a descriptive analysis and a predictive model are performed. Significant relations between some socio-demographic variables and the assessed blood pressure are obtained. The present work describes the statistical approach estimating a model for relevant information of questionnaire by Generalized Linear Models. This approach is still going on.

Keywords: Childhood, Hypertension, Caregivers, General Linear Models.

A New Generalized Class of Bivariate Distributions based on Latent Random Variables

Manuel Franco[1], Juana-Maria Vivo[1], Debasis Kundu[2]
[1]Dept of Statistics and Operations Research, University of Murcia, Spain, [2]Dept of Mathematics and Statistics, Indian Institute of Technology, India

Bivariate lifetime data arise in many fields such as medicine, biology, public health, epidemiology, engineering, economic and demography, being very important to considerer different bivariate distributions that could be used to model such bivariate lifetime data as well as their properties are also useful to carry out such a purpose.

One of the most cited lifetime models is the Marshall-Olkin bivariate exponential (MOBE) distribution introduced in 1967 by Marshall and Olkin, based on a latent factor through the minimization process, which is motivated by a competing risks model and a shock model. As it can be found in the literature, this bivariate lifetime distribution along with its generalizations have played an important role in life testing, reliability, survival and other fields of applications. Some modifications and extensions of the MOBE have also been studied, e.g. see Sarhan and Balakrishnan (2007), Franco and Vivo (2010), Kundu and Gupta (2010a) and Franco, Kundu and Vivo (2011) and the references therein. A recent more generalized bivariate distribution family is the generalized Marshall-Olkin bivariate distribution (GMOB) class introduced by Gupta, Kirmani and Balakrishnan (2013).
In this work, we propose a new generalized class of bivariate distribution models based on the maximization process introduced by Kundu and Gupta (2009), also used by Kundu and Gupta (2010b) and Kundu, Franco and Vivo (2014), which is motivated by a stress model and a maintenance model. This generalized class contains as particular cases the bivariate generalized exponential model of Kundu and Gupta (2009), the bivariate proportional reversed hazard rate model of Kundu and Gupta (2010b), the bivariate log-exponentiated Kumaraswamy model of Elsherpieny et al. (2014) and the bivariate exponentiated modified Weibull extension of El-Gohary et al. (2016).

Keywords: Latent random variable, bivariate distribution, Marshall-Olkin model, competing risk model.

An SEM Approach to Modelling Housing Values

Jim Freeman, Xin Zhao
Alliance Manchester Business School,
University of Manchester, United Kingdom

Though hedonic regression remains a popular technique for estimating property values, structural equation modeling (SEM) is increasingly seen as a realistic analysis alternative. The article presents an SEM analysis of a historical dataset for a large Canadian realtor. An iterative approach was adopted for the modelling, the first phase focusing on internal relationships between houses' structural characteristics and the second, on housing values and their determinants. In the final phase, advertised list prices and location details were the priority. A comprehensive evaluation of the resulting holistic model revealed a wealth of significant structural relationships - particularly between House Style, House Structure and House Attributes.

Keywords: AMOS, Hedonic Regression, Housing values, SEM.

Rates of Approximation of Integral Functionals of Markov Processes with Applications

Iurii Ganychenko
Department of Probability Theory, University of Potsdam, Germany

We provide weak and strong rates of approximation of integral functionals of Markov processes by Riemann sums. Assumptions on the processes are formulated only in terms of their transition probability densities and therefore are quite flexible. Namely, we pose a proper boundary condition on the derivative of the transition probability density of the respective Markov process with respect to the time variable. The class of processes under consideration includes diffusion processes, stable processes and models with Lévy(-type) noises.
We focus on integral functionals with non-regular kernels. As a particular important example of such a kernel, we consider an indicator function and the occupation time of a Markov process as a respective integral functional. We apply the results of weak and strong approximation rates of integral functionals to the estimates of the error of approximation of the price of an occupation time option.

Keywords: Integral functional, Rates of approximation, Markov process, Occupation time option.

Comparison of Stochastic Processes

Jesús E. García, V.A. González-López
University of Campinas, Brazil

In this paper is introduced a distance which allows to compare Markovian processes. It is shown the relationship of this distance to the divergence of Kullback Leibler and revealed its stochastic behavior in terms of the Chi-squared distribution. The distance allows to decide if there is any discrepancy between two samples of stochastic processes. When a discrepancy exists, the use of this distance allows us to find the strings where the discrepancy is manifested. We apply the distance to written texts of European Portuguese coming from two authors: Vieira-1608 and Garrett-1799. In the application the distance reveals the linguistic configurations that expose discrepancies between written texts of different genres from the same author. This type of results could characterize linguistic genres and varieties in the same language.

Keywords: Distance, Partition Markov Models, Kullback Leibler, Chi-square distribution, Computational linguistics.

References:

[1]. Galves, A.; Galves, C.; García, J.E.; Garcia, N.L.; Leonardi, F. Context tree selection and linguistic rhythm retrieval from written texts. The Annals of Applied Statistics 2012, 186-209.
[2]. García, J.E.; González-López, V.A.; Kubo de Andrade, F.H. Dissimilarity between Markovian Processes applied to Industrial Processes. In AIP Conference Proceedings (forthcoming).
[3]. García, J.E.; González-López, V.A. Detecting regime changes in Markov models. In SMTDA2014 Book.
[4]. García, J.E.; González-López, V.A. Consistent Estimation of Partition Markov Models. Entropy 2017, *19*, 160.

Stochastic Distance between Burkitt lymphoma/leukemia Strains

Jesús E. García, V. A. González-López
University of Campinas, Brazil

Quantifying the proximity between N-grams allows to establish criteria of comparison between them. Recently, a consistent criteria (d) to achieve this end, was proposed, see [1]. This criterion takes advantage of a model structure on Markovian processes in finite alphabets and with finite memories, called Partition Markov Models, see [2]. It is possible to show that d goes to zero almost surely, when the compared processes follow identical law and the samples sizes grow. In this work we explore the performance of d in a real problem, using d to establish a notion of natural proximity between DNA sequences from patients Burkitt lymphoma. And we present a robust strategy of estimation to identify the law that governs most of the sequences considered, thus mapping out a common profile to all these patients, via their DNA sequences.

Keywords: Partition Markov Models, Bayesian Information Criterion, Robust Estimation in Stochastic Processes.

References:

[1]. García, J.E.; González-López, V.A. Detecting regime changes in Markov models. In SMTDA2014 Book.
[2]. García, J.E.; González-López, V.A. Consistent Estimation of Partition Markov Models. Entropy 2017, 19, 160.

Modelling Dietary Exposure to Chemical Components in Heat-Processed Meats

Stylianos Georgiadis[1,3], Lea Sletting Jakobsen[2], Bo Friis Nielsen[1,3], Anders Stockmarr[1], Elena Boriani[2,3], Lene Duedahl-Olesen[2,] Tine Hald[2,3], Sara Monteiro Pires[2]

[1]Dept of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark,
[2]The National Food Institute, Technical University of Denmark, Denmark,
[3]Global Decision Support Initiative, Technical University of Denmark, Denmark

Several chemical compounds that potentially increase the risk of developing cancer in humans are formed during heat processing of meat. Estimating the overall health impact of these compounds in the population requires accurate estimation of the exposure to the chemicals, as well as the probability that different levels of exposure result in disease. The overall goal of this study was to evaluate the impact of variability of exposure patterns and uncertainty of exposure data in burden of disease estimates. We focus on the first phase of burden of disease modelling, i.e. the estimation of exposure to selected compounds in the Danish population, based on concentration and consumption data. One of the challenges that arises in the probabilistic modelling of exposure is the presence of "artificial" zero counts in concentration data due to the detection level of the applied tests. Zero-inflated models, e.g. the Poisson-Lognormal approach, are promising tools to address this obstacle. The exposure estimates can then be applied to dose-response models to quantify the cancer risk.

Keywords: Burden of disease, Exposure modelling, Model fitting.

Utilizing Customer Requirements' Data to Link Quality Management and Services Marketing Objectives

Andreas C. Georgiou, Kamvysi Konstantina, Gotzamani Katerina, Andronikidis Andreas
Dept of Business Administration, University of Macedonia, Greece

This study is built upon the argument that when organizations develop a service strategy, their management should acknowledge the significance of the Voice of the Customer in defining service quality, and interpret it into a set of prioritized marketing strategies to guide service design activities. In this respect, the aim of this work is twofold: (a) to propose and implement a Quality Function Deployment (QFD) framework, comprising a 3-phased process for planning service strategy grounded to customers' data reflecting their requirements, thus aligning quality

management and services marketing and linking decisions related to market segmentation, positioning, and marketing mix (b) to employ the QFD method in the enhanced environment of the linear programming method of LP-GW-Fuzzy-AHP (Kamvysi, Gotzamani, Andronikidis and Georgiou, 2014) in order to capture and prioritize uncertain and subjective judgments, which reflect the true "Voice of the Customer". The proposed QFD framework is implemented in the banking sector for planning service marketing strategies.

Keywords: Voice of the Customer, Quality Management, Services Marketing, QFD, Fuzzy AHP.

Efficiency Evaluation of Multiple-Choice Exam

Evgeny Gershikov, Samuel Kosolapov
ORT Braude Academic College of Engineering, Israel

Multiple-Choice Exams are widely used in colleges and universities. Simple forms filled by students are easy to check and the forms can even be graded automatically using a scanner or camera-based systems that utilize image processing and computer vision techniques. These techniques are used to align the answer sheets, segment them into the relevant regions and read the answers marked by the students. Then the grades can be easily calculated by comparing the marked data with the correct answers. However, once the grades have been derived, it is of interest to analyze the performance of the students in this particular exam and compare it to other groups of students or past examinations. In addition to the basic statistical analysis of calculating the average, the standard deviation, the median, the histogram of the grades, the passing/failing percent of students and other similar values, we propose efficiency measures for each question and for the whole exam. One of these efficiency measures attempts to answer the following question: how many of the "good" students have answered a particular question correctly. Another measure attempts to evaluate the performance of the "bad" students: how many of them have failed in a particular question. A question is considered efficient if most "good" students succeed in it while most "bad" ones fail. In a similar fashion, an exam questionnaire is considered efficient if the majority of its questions are efficient. Our measures can be used both for multiple-choice and numeric answers (where points are granted if the student writes the expected numeric value or one close to it). We have performed the proposed statistical analysis on the grades of a number of real life examinations. Our conclusion is that the proposed analysis and efficiency measures are beneficial for the purpose of estimating the quality of the exam and

locating it weakest links: the questions that fail to separate the “good” and the “bad” students.

Keywords: Multiple-Choice Exam, efficiency measure, statistical analysis.

Topic detection using the DBSCAN-Martingale and the Time Operator

Ilias Gialampoukidis[1;2], Stefanos Vrochidis[2], Ioannis Kompatsiaris[2], Ioannis Antoniou[1]

[1]Dept of Mathematics, Aristotle University of Thessaloniki, Greece, [2]Information Technologies Institute, Centre for Research and Technology-Hellas, Greece

Topic detection is usually considered as a decision process implemented in some relevant context, for example clustering. In this case, clusters correspond to topics that should be identified. Density-based clustering, for example, uses only a density level and a lower bound for the number of points in a cluster. As the density level is hard to be estimated, a stochastic process, called the DBSCAN-Martingale, is constructed for the combination of several outputs of DBSCAN for various randomly selected values of ϵ in a predefined closed interval $[0, \epsilon_{max}]$ from the uniform distribution. We have observed that most of the clusters are extracted in the interval $[0, \epsilon_{max}/2]$, and moreover in the interval $[\epsilon_{max}/2\,, \epsilon_{max}]$ the DBSCAN-Martingale stochastic process is less innovative, i.e. extracts only a few or no clusters. Therefore, non-symmetric skewed distributions are needed to generate density levels for the extraction of all clusters in a fast way. In this work we show that skewed distributions may be used instead of the uniform, so as to extract all clusters as quickly as possible. Experiments on real datasets show that the average innovation time of the DBSCAN-Martingale stochastic process is reduced when skewed distributions are employed, so less time is needed to extract all clusters.

Keywords: DBSCAN-Martingale, Time Operator, Skewed distributions, Internal Age, Density-based Clustering, Innovation process.

References:

[1]. L. Devroye. Non-uniform random variate generation, Springer, New York, 1986.

[2]. M. Ester, H. P. Kriegel, J. Sander and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 96, 34, 226{231, 1996.

[3]. I. Gialampoukidis and I. Antoniou. Time Operator and Innovation. Applications to Financial Data. In Proceedings of the 3rd Stochastic Modelling Techniques and Data Analysis, Lisbon, 2014, pp. 269{281

[4]. I. Gialampoukidis, K. Gustafson and I. Antoniou. Financial Time Operator for random walk markets. Chaos, Solitons & Fractals, 57, 62{72, 2013.

[5]. I. Gialampoukidis, S. Vrochidis and I. Kompatsiaris. A Hybrid Framework for News Clustering Based on the DBSCAN-Martingale and LDA. In Machine Learning and Data Mining in Pattern Recognition, Springer International Publishing, 170{184, 2016.

Methodological Issues in the Three-Way Decomposition of Mortality Data

Giuseppe Giordano[1], Steven Haberman[2], Maria Russolillo[1]
[1]*Dept of Economics and Statistics, University of Salerno, Italy,*
[2]*Cass Business School, City University London, United Kingdom*

The three-way model has been proposed as an extension of the original Lee-Carter (LC) model when a three mode data structure is available. The three-way LC model allows to enrich the basic LC model by introducing several tools of exploratory data analysis. Such exploratory tools allow to give a new perspective to the demographic analysis supporting the analytical results with a geometrical interpretation and a graphical representation. From a methodological point of view there are several issues to deal with when focusing on such kind of data. Specially, in presence of the three way data structure, there are several choices on data pre-treatment that will affect the whole data modelling.
The first step of three-way mortality data investigation should be addressed by exploring the different source of variations and highlighting the significant ones. We consider the three-way LC model investigated trough a three-way analysis of variance with fixed effects, where each cell is given by the mortality rate in a given year of a specific age-group for a country. We can thus analyze the 3 main effects, the 3 two-way interactions and 1 three-way interaction.
In this paper we propose to consider the death rates aggregated for time, age-groups and countries. First we consider the variability attached to the three ways: Age, years and countries. Furthermore, we may consider the variability induced by the paired interactions between the three ways. Finally, the three way interaction could give information on which country have a specific trend (along years) in each age-group. This kind of analysis is recommended to assess the source of variation in the raw mortality data, before to extract rank-one components.

Keywords: Anova, Lee-Carter Model, Three-way principal component analysis.

Linear Approximation of Nonlinear Threshold Models

Francesco Giordano, Marcella Niglio, Cosimo D. Vitale
Dept of Economics and Statistics,
Università degli Studi di Salerno, Italy

The complexity of most nonlinear models often leads to evaluate if a linear representation can be admitted for this class of models in order to take advantage of the large and strengthened literature developed in the linear domain. Unfortunately linear representation of nonlinear models have been obtained only in few and well defined cases (see Bollerslev (Journal of Econometrics, vol. 31, 307-327, 1986), Gourieroux and Montfort (Journal of Econometrics, vol. 52, 159-199, 1992), among the others).
The aim of our contribution is to define "the best" linear approximation of the nonlinear Self Exciting Threshold Autoregressive (SETAR) model (Tong and Lim, Journal of the Royal Statistical Society (B), vol. 42, 245-292, 1980): in more detail, differently from the cited literature, our aim is to find a theoretically best linear approximation such that the nonlinear SETAR process $\{Y_t\}$ can be decomposed as:

$$Y_t=W_t+X_t$$

where W_t is the linear approximation of Y_t and X_t is the remaining nonlinear component. Moreover, we show that X_t is again a SETAR process with some restrictions on its parameters.
This decomposition has at least two main advantages:

- in model selection it allows to proper discriminate between linear and nonlinear structures using a theoretically approach to derive W_t ;
- the investigation of the "purely'" nonlinear component X_t can be used to identify nonlinear features for the SETAR process.

Keywords: Nonlinear SETAR process, linear approximation.

Some Remarks on the Prendiville Model in the Presence of Catastrophes

Virginia Giorno[1], Serena Spina[2]
{[1]Dipartimento di Informatica, [2]Dipartimento di Matematica,}
Università di Salerno, Italy

In the last four decades, time-continuous Markov-chains have been extensively studied under the effect of random catastrophes. Specifically, a catastrophe is an event occurring at random times and it produces an instantaneous variation of the state of the system by passing from the current state to a specified state that can be zero. Catastrophes play a relevant role in various contexts. For example, a catastrophe at zero

state can be considered as the effect of a fault that clears the queue, while in a population dynamics a catastrophe can be interpreted as the effect of an epidemic or an extreme natural disaster (forest fire, flood, …).
Among other processes, the logistic process (proposed in 1949 by Prendiville) plays a considerable role because it was always used in a variety of biological and ecological contexts due to its versatility and its mathematical tractability. The Prendiville process is a continuous-time Markov-chain defined on a finite space and characterized by state dependent rates: the births are favorite when the population size is low and the deaths are more frequent for large size populations. In the present paper we focus on the non-homogeneous logistic process in the presence of catastrophes, by interpreting it in the contest of the queue theory.
We analyze the effect of the catastrophes on the state of the system and the first crossing time.

Keywords: Markov-chains, Catastrophes, Prendiville process.

A Multi-Mode Model for Stochastic Project Scheduling with Adaptive Policies based on the Starting Times and Project State

Pedro Godinho
CeBER and Faculty of Economics, University of Coimbra, Portugal

This paper presents a model for multi-mode stochastic project scheduling in which there is a due date for concluding the project and a tardiness penalty for failing to meet this due date. Several different modes can be used to undertake each activity, some slower and less expensive and others faster but more expensive. The mode used for undertaking each activity is chosen immediately before the activity starts, and such choice is based on an adaptive policy: a set of rules that defines the execution mode according to the way the project is developing.
Building on previous work, we use the starting times of the activities to define the policies: we define a set of thresholds, and the starting time of the activity is compared with those thresholds in order to define the execution mode. Differently from previous work, we also include an indicator of the global project state at the start of the activity in the rules that define the execution mode. The rationale is that, if the project as a whole is very late and the activity is starting somewhat late, then the probability of the activity being critical is smaller, and there is also a smaller incentive to use a faster, more expensive mode.

We use an electromagnetism-like heuristic for choosing a scheduling policy. We apply the model to a set of projects previously used by other authors, and we compare the results with and without including the project state in the policy definition. We discuss the characteristics of the cases in which the incorporation of the project state leads to larger gains.

Keywords: Stochastic project scheduling, Metaheuristics, Simulation.

An Application of Data Mining Methods to the Analysis of Bank Customer Profitability and Buying Behavior

Pedro Godinho[1], Joana Dias[2], Pedro Torres[3]
[1]CeBER and Faculty of Economics, University of Coimbra, Portugal, [2]CeBER, INESC-Coimbra and Faculty of Economics, University of Coimbra, Portugal, [3]CeBER and Faculty of Economics, University of Coimbra, Portugal

In this paper we use a database concerning the behavior of customers from a Portuguese bank to analyse churn, total wealth deposited in the bank, profitability and next product to buy. The database includes data from more than 94000 customers, and includes all transactions and balances of bank products from those customers for the year 2015. We describe the main difficulties found concerning the database, as well as the initial filtering and data processing necessary for the analysis. We discuss the definition of churn criteria and the results obtained by the application of several data mining techniques for churn prediction and for the short term forecast of future profitability. We present the results of a clustering analysis of the main factors that determine client profitability and client wealth deposited in the bank. Finally, we present a data mining- based model for predicting the next product that will be bought by a client. The models show some ability to predict churn, but the fact that the data concerns just a year clearly hampers their performance. In the case of the forecast of future profitability, the results are also hampered by the short time frame of the data and by some outliers present in the data. The clustering analysis shows that age is the most important factor in determining total wealth deposited in the bank and customer profitability, followed by the number of bank card transactions and the number of logins to the bank site (in the case of profitability) and stock market transactions (in the case of total wealth deposited in the bank). The models for the next product to buy show a very encouraging performance, being able to achieve a good detection ability for the main products of the bank.

Keywords: Data Mining, Bank Marketing, Churn, Clustering, Random Forests.

Piece-wise Quadratic Approximations of Subquadratic Error Functions for Machine Learning

A. N. Gorban[1], E. M. Mirkes[1], A. Zinovyev[2]
[1]*Dept of Mathematics, University of Leicester, United Kingdom,*
[2]*Institut Curie, PSL Research University, France*

Most of machine learning approaches have stemmed from the application of minimizing the mean squared distance principle, based on the computationally efficient quadratic optimization methods. However, when faced with high-dimensional and noisy data, the quadratic error functionals demonstrated many weaknesses including high sensitivity to contaminating factors and dimensionality curse. Therefore, a lot of recent applications in machine learning exploited properties of non-quadratic error functionals based on L1 norm or even sub-linear potentials corresponding to quasinorms Lp ($0<p<1$). The back side of these approaches is increase in computational cost for optimization.

We develop a new machine learning framework (theory and application) allowing one to deal with arbitrary error potentials of not-faster than quadratic growth, imitated by piece-wise quadratic function of subquadratic growth (PQSQ error potential). We elaborate methods for constructing the standard data approximators (mean value, k-means clustering, principal components, principal graphs) for arbitrary non-quadratic approximation error with subquadratic growth and regularized linear regression with arbitrary subquadratic penalty by using a piecewise-quadratic error functional (PQSQ potential). These problems can be solved by expectation-minimization algorithms, which are organized as solutions of sequences of linear problems by standard and computationally efficient methods.

The suggested methodology has several advantages over existing ones:

(a) Scalability: the algorithms are computationally efficient and can be applied to large data sets containing millions of numerical values.

(b) Flexibility: the algorithms can be adapted to any type of data metrics with subquadratic growth, even if the metrics cannot be expressed in explicit form. For example, the error potential can be chosen as adaptive metrics.

(c) Built-in (trimmed) robustness: choice of intervals in PQSQ can be done in the way to achieve a trimmed version of the standard data approximators, when points distant from the approximator do not affect the error minimization during the current optimization step.

(d) Guaranteed convergence: the suggested algorithms converge to local or global minimum just as the corresponding predecessor algorithms based on quadratic optimization and expectation/minimization-based splitting approach.

Further details and references could be found in e-print [1].

Keywords: machine learning; expectation-minimization; lasso; (min,+) algebra.

Reference: A.N. Gorban, E.M. Mirkes, A. Zinovyev, Piece-wise quadratic approximations of arbitrary error functions for fast and robust machine learning, arXiv:1605.06276 [cs. LG].

Blessing of Dimensionality: One-Trial Correction of Legacy AI Systems in High Dimension

A. N. Gorban[1], I. Romanenko[2], R. Burton[1,2], I. Y. Tyukin[1]
[1]*Dept of Mathematics, University of Leicester, United Kingdom,*
[2]*Imaging and Vision Group, ARM Holding, United Kingdom*

Despite their success and utility, legacy AI systems occasionally make mistakes. Their generalization errors may be caused by many various issues. We propose that the legacy AI system itself be augmented by miniature low-cost and low-risk additions. These additions are, in their essence, small neural network cascades. Such small neuronal cascades can be constructed via simple noniterative procedures for a large class of legacy AI systems operating with high-dimensional data, We prove this by showing that in an essentially high-dimensional finite random set with probability close to one all points are extreme, i.e. every point is linearly separable from the rest of the set. Such a separability holds even for large random sets up to some upper limit, which grows exponentially with dimension. Thus, in high-dimensional data space a single element, i.e. a mistake, can be separated from the rest by a simple perceptron.

The most inspiring consequence of the measure concentration phenomena is the paradigm shift. It is a common point of view that the complex learning systems should produce complex knowledge and skills. On contrary, it seems to be possible that the main function of many learning system, both technical and biological, in addition to production of simple skills, is a special preprocessing. They transform the input flux ('reality') into essentially multidimensional and quasi-random distribution of signals and images plus, may be, some simple low dimensional and more regular signal. After such a transformation, ensembles of non-interacting or weakly interacting small neural networks ('correctors' of simple skills) can solve complicated problems.

For computational testing of the algorithms, we used the original training and testing sets as well as three different videos (not used in the training set generation) to test trash model creation and its effectiveness at removing false positives. As a legacy AI system we selected a Convolutional Neural Network trained to detect objects in images.

Further details and references could be found in e-print [1].

Keywords: big data, measure concentration, machine learning, visual intelligence, convolutional neural network.

Reference: A.N. Gorban, I. Romanenko, R. Burton, I.Y. Tyukin, One-Trial Correction of Legacy AI Systems and Stochastic Separation Theorems, arXiv:1610.00494 [stat.ML].

Embedded Semi-Markov Process as Reliability Model of Two Different Units Renewal Cold Standby System

Franciszek Grabski
Dept of Mathematics and Physics, Polish Naval University, POLAND

An embedded semi-Markov stochastic process is applied in reliability problem. The problem concerns of two different units renewal cold standby system. We assume that the system consists of one operating unit *A*, the stand-by unit *B* that may have different probability distributions of the times to failure. We suppose that there is an unreliable switch in the system which is used at the moment of the working unit failure. A discrete state space and continuous time stochastic process describes work of the system in reliability aspect. To obtain the reliability characteristic and parameters of the system we construct so called an embedded semi-Markov process in this process. In our model the time to failure of the system is represented by a random variable denoting the first passage time from the given state to the subset of states. To calculate the reliability function and mean time to failure of the system we apply theorems of the Semi-Markov processes theory concerning a probability distribution of a first passage time to a subset of states for semi-Markov process. Often an exact reliability function of the system by using Laplace transform is difficult to calculate, frequently impossible. In those cases we can apply one of theorems of Semi-Markov processes perturbation theory, to obtain an approximate reliability function of the system.

Keywords: Semi-Markov process, cold standby system, embedded stochastic process.

Towards Prediction of Catastrophic Failure Events of Laser-Induced Damage in Optical Laser Elements

Povilas Grigas[1], Rūta Levulienė[2], Vilijandas Bagdonavičius[2], Andrius Melninkaitis[1]
[1]Laser Research Center, Vilnius University, Lithuania,
[2]Dept of Mathematical Statistics, University of Vilnius, Lithuania

Every laser is a chain of optical elements, that could be damaged by the self-generated intense light. Thus question “how long laser will work the until its self-damage?” is very important for everyone who is using lasers. The answer to this question directly related to both lifetime of optical parts and statistical properties of generated light. To quantify laser damage performance of optical element as well as predict its failure free operation time nowadays damage frequency method is used together with various phenomenological regression models. To our best knowledge survival analysis methods are not widely used to study above mentioned problems. In this study, we explore potential of survival analysis by directly comparing it with “classical” damage frequency method. Firstly, we introduce the context and define simplified concept of Laser-Induced Damage Threshold (LIDT) testing problem when pulsed lasers are used. To compare benefits and shortcomings of survival analysis and damage frequency method real experimental data are used. Furthermore, to check the validity of both methods a Monte Carlo simulations were performed in well controlled manner. By doing so we were able to make new insights about the applicability and accuracy of both statistical methods for characterization of damage threshold processes in optics.

Keywords: laser-induced damage, accelerated failure time model, damage frequency, LIDT, AFT.

An Interest-Rate Model with an Unobservable Mean-Reversion Level

Stefanie Grimm[1], Christina Erlwein-Sayer[1], Rogemar Mamon[2]
[1]Dept of Financial Mathematics, Fraunhofer Institute for Industrial Mathematics ITWM, Germany, [2]Dept of Statistical and Actuarial Sciences, University of Western Ontario, Canada

We consider a regime-switching model for the short-term interest-rate related to the well-known Vasicek model. One of our basic assumptions thereby is that the switches in regime are not triggered by distinguishable single events. Rather, they result from a multitude of incidents, which combined build the underlying mood of the market. To reflect these

dynamics in a mathematically reasonable way, we choose an Ornstein-Uhlenbeck process involving a mean-reversion level that is guided by a continuous-time, finite-state Markov chain. Additionally, we assume partial observation. That is, we assume an observable short-rate but unobservable switches in regime. We consider recursive, finite-dimensional filters for the Markov chain and related processes employing a change to an idealized measure. Based on that, by using the expectation maximization algorithm, we obtain on-line parameter estimates. We discuss an application of the algorithm to daily German treasury bill data and compare those results to a set of foreign treasury bills.

Keywords: Regime-Switching, Hidden Markov Model, Interest-Rate.

Limit Theorems for Queueing Systems with Different Service Disciplines

Svetlana Grishunina
Dept of Probability Theory, Lomonosov Moscow State University; Moscow Institute of Electronics and Mathematics, National Research University Higher School of Economics, Russia

In this paper we investigate a multi-server queueing system with regenerative input flow and independent service times with finite mean. Queues with several servers are sufficiently complex but considerably interesting. There are many papers devoted to this theme. We consider queueing systems with various rules (disciplines) of the service performance: systems with a common queue and systems with individual queues in front of the servers. In the second case an arrived customer chooses one of the servers in accordance to a certain rule and stays in the chosen queue up to the moment of its departure from the system. We define some classes of disciplines and analyze the asymptotical behavior of a multi-server queueing system in a heavy-traffic situation (traffic rate $\rho > 1$). The main result of this work is the weak convergence of scaled processes of waiting time and queue length to the process of the Brownian motion.

Keywords: Queueing System, Heavy-traffic, Limit Theorems, Service Disciplines.

Acknowledgment: *The research was partially supported by RFBR grant 17-01-00468.*

Fractional Derivative of the Gabor-Morlet wavelet – an Application in Volcanology

E. Guariglia[1,2], G. Pucciarelli[1], S. Silvestrov[2]
[1]*Dept. of Physics "E. R. Caianiello", University of Salerno, Italy,*
[2]*Division of Applied Mathematics, Mälardalen University, Sweden*

A fractional-wavelet analysis of volcanological data is presented. In the first part the Gabor-Morlet wavelet is used, while in the second one, in order to obtain more precious information, the chosen mother wavelet is its fractional derivative. Both methods show occurrence of a particular type of tides (seiches) but the second method provides more details.

Keywords: Fractional derivative, Gabor-Morlet wavelet, continuous wavelet transform, volcanology, seiches.

Stepwise Regression – an Application in Earthquakes Localization

E. Guariglia[1,2], G. Pucciarelli[1], S. Silvestrov[2]
[1]*Dept. of Physics "E. R. Caianiello", University of Salerno, Italy,*
[2]*Division of Applied Mathematics, Mälardalen University, Sweden*

In this paper, the multiple linear stepwise regression is applied to the earthquakes localization. The stepwise model empathizes how it contributes to find a solution in earthquakes localization, by describing its conditions of use in a software for the computation of seismic sources' collocation (HYPO71PC). The stepwise regression allows us to reach a balance between the number of independent variables which we should use and their capacity of description. In particular, the case of Mount Vesuvius (south of Italy) is widely examined and discussed.

Keywords: Stepwise regression, earthquake localization, variance, HYPO71PC.

A Spectral Analysis of the Weierstrass-Mandelbrot Function on the Cantor Set

Emanuel Guariglia[1], Xiaomin Qi[2], Sergei Silvestrov[3]
[1]*Dept of Physics "E. R. Caianiello", University of Salerno, Italy,*
[2]*School of Education, Culture & Communication, Mälardalen University, Sweden,*
[3]*School of Education, Culture & Communication, Mälardalen University, Sweden*

In this paper, the Weierstrass-Mandelbrot function on the Cantor set is presented with emphasis on possible applications in science and

engineering. An asymptotic estimation of its one-said Fourier transform, in accordance with the simulation results, is analytically derived. Moreover, a time-frequency analysis of the Weierstrass-Mandelbrot function is provided by numerical computation of its continuous wavelet transform.

Keywords: Weierstrass-Mandelbrot function, Cantor set, one-said Fourier transform, continuous wavelet transform.

Some Asymptotic Results for Truncated-Censored and Associated Data

Zohra Guessoum, Abdelkader Tatachak
MSTD Laboratory, Dept of Probability and Statistic,
University of Sciences and Technology Houari Boumediene, Algeria

Left truncation and right censoring (LTRC) arise frequently in practice for life data. Under LTRC model, the product limit estimator (PLE) was proposed and investigated in the i.i.d. case by Tsai et al. (1987). In the presence of covariates, the conditional version was studied in the α - mixing setting by Liang et al. (2012). Our focus in the present paper is to assess strong uniform consistency rates for the cumulative hazard and the product limit estimates when the lifetime observations form an associated sequence. Then, as an application we derive a strong uniform consistency rate for the kernel estimator of the hazard rate function introduced and studied in the i.i.d. case by Uzunogullari and Wang (1992). Some simulations are drawn to support our theoretical results.

Keywords: Associated data, Left truncation, Right censoring, Strong uniform consistency rate.

References:

[1]. Liang HY, de Uña-Álvarez J, Iglesias-Pérez MdC (2012) Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. Test. 21:790-810.

[2]. Tsai WY, Jewell NP, Wang MC (1987) A note on the product-limit estimator under right censoring and left truncation. Biometrika. 74:883--886.

[3]. Uzunogullari Ü, Wang JL (1992) A comparison of hazard rate estimators for left truncated and right censored data. Biometrika. 79:297--310.

Modelling Spread Limitations of Oil Spills at Sea

Sambor Guze, Krzysztof Kołowrocki, Jolanta Mazurek
Dept of Mathematics, Faculty of Navigation, Gdynia Maritime University

To describe the oil spill central point position a two-dimensional stochastic process is used and its drift trend curve is determined. The oil spill domain movement general model for various hydro-meteorological conditions is constructed and the method of this model unknown parameters estimation is proposed. These methods are used to predict the spill domain movement and to prevent and to mitigate the oil spill consequences by constructing the algorithm for oil spill spread limitations. An exemplary application of this procedure is given.

Keywords: drift trend, oil spill domain, spread limitations, stochastic model.

Acknowledgments: *Paper in the scope of HAZARD project funded from the Interreg Baltic Sea Region Programme 2014-2020 under grant Agreement No #R023.*

Longevity Trends and their Impact on Life Expectancy and Annuity Values – how Fast are they Changing?

Steven Haberman, Zoltan Butt
Cass Business School, City University London, United Kingdom

There is considerable empirical evidence that mortality rates are falling across the age range in many countries, leading to the widely discussed phenomenon of longevity risk. Researchers have proposed a range of models to represent this downward secular trend – many models are based on the assumption of an exponential decline. See for example, the models of Lee and Carter (1992), Benjamin and Soliman (2000), Sithole et al (2000), Renshaw and Haberman (2003) and Haberman and Renshaw (2012). In this exploratory piece of work, we consider the effect of this trend on two important summary indices by considering their time derivatives – the indices are the expectation of life at age x, widely used by demographers to summarise a set of mortality rates, and the discounted annuity value, widely used by insurance companies and pension plans to assess the value of future cash flows. We consider the period and cohort versions of these indices and consider the relationship between them and between their time derivatives. In this regard, we build on the work of Vaupel (1986), Schoen and Canudas-Romo (2005), Haberman et al (2011) and Missov and Lenart (2011).

Keywords: Longevity risk, Cohort life expectancy, Discounted annuity value.

Fitting Markovian Binary Trees using Global and Individual Demographic Data

Sophie Hautphenne, Katharine Turner, Melanie Massaro
The University of Melbourne & EPFL, Australia & Switzerland

We estimate the parameters of the transient Markovian arrival process (TMAP) controlling the individuals lifetime and reproduction epochs in a Markovian binary tree. The datasets used are population data containing information on age-specific fertility and mortality rates, and we apply a non-linear regression method or a maximum likelihood method, depending on the precision of the available data. We discuss the optimal choice of the number of phases in the TMAP, and we provide confidence intervals for the model outputs. The results are then applied using real data on the endangered black robin bird population.

Cluster Validation: How to Think and What to do?

Christian Hennig
Dept of Statistical Science, UCL, United Kingdom

Cluster analysis is about finding groups in data. There are many cluster analysis methods and on most datasets clusterings from different methods will not agree. Cluster validation concerns the evaluation of the quality of a clustering. This is often used for comparing different clusterings on a dataset, stemming from different methods or with different parameters such as the number of clusters.
There are many aspects of cluster validity. Some of these aspects are mostly informal, such as the question whether a clustering makes substantive sense, and the visual evaluation of a clustering. There are also various measurements for cluster validity. Often these are used in such a way that the validity of the whole clustering is measured by a single number. But the quality of a clustering is rather multivariate; within-cluster homogeneity, between-cluster separation, representation of cluster members by a centroid object or stability could be measured, and what is most important depends on the aim of clustering.
In this presentation I will give an overview of techniques for cluster validation particularly focusing on a number of new measurements of different aspects of cluster validity. I will also discuss the issue what the "true clusters" are that we want to find and how this depends on the specific application and the aims and concepts of the researcher, so that these can be connected to specific techniques for cluster validation.

Keywords: cluster analysis, homogeneity, separation, cluster quality.

A Simple Test of Monotonicity and Monotonicity-Related Properties

Javier Hidalgo, Tatiana Komarova
London School of Economics and Political Science, United Kingdom

We develop a test for monotonicity in a nonparametric framework using partial sums empirical process. We show that the test has suitable asymptotic properties. In particular we show after appropriate transformation the asymptotic distribution is a functional of a standard Brownian motion, so that critical values are available. However, due to the possible poor approximation of the asymptotic critical values to the finite sample ones, we also describe a valid bootstrap algorithm. We show how methodology can be extended to test for other properties of the regression function such as convexity, concavity, absolute monotonicity and U-shape. We outline how this can extended to a framework when other covariates are present and no monotonicity-related properties are imposed on those.
We also establish how monotonicity can be tested in the situation of endogeneity if there is a strong instrument available. We outline some applications in economics and finance.

Keywords: Monotonicity, convexity, concavity, U-shape. Distibution-free-estimation.

Importance of Factors Contributing to Work-Related Stress: Comparison of Four Metrics

Mounia N. Hocine[1], Natalia Feropontova[2], Ndèye Niang[3], Karim Aït-Bouziad[1], Gilbert Saporta[3]
[1]Modelisation, Epidemiology and Surveillance of Health-related Risks team, Conservatoire National des Arts et Métiers (CNAM), France, [2]Faculty of Applied Mathematics and Cybernetics, Tomsk State University, Russia, [3]Centre d'Etude et De Recherche en Informatique et Communications, France

In the field of management research, decision makers would like to be provided with statistical tools that can help them identify risk factors requiring priority action to achieve desirable outcomes such as reducing work-related stress levels. The aim is to identify the best drivers of improvement, and quantify their respective impacts. However, as predictors are often correlated, regression coefficients cannot be used directly to provide decision makers with ranked predictors. To overcome this limit, the Weifila method has been proposed, which is based on variance decomposition in the linear regression context.

Here, we hierarchize risk factors in terms of their impact on the outcome of interest, using four different metrics. The first is based on the Weifila method, the second on random forests, the third on attributable risk (an epidemiological indicator), and the fourth on path coefficients in a PLS-SEM model.
This study was motivated a large work-related stress level dataset with 10,000 anonymized employees who completed two questionnaires in a face-to-face interview with an occupational physician. The first, on 25 stress-related items, was subsequently used to build a stress scale (the outcome of interest). The second questionnaire involved 58 psychosocial risk factors on a 6-points Likert scale.
The results show similar rankings for the ten first items for the four different metrics. The attributable risk is the easiest tool to use for managers, but requires a causal assumption that needs further analysis.

Keywords: Importance, Weifila, random forest, attributable risk, regression, ranking, occupational stress, PLS path modeling.

Hidden Variable Models for Market Basket Data

Harald Hruschka
Dept of Marketing, University of Regensburg, Germany

We compare the performance of several hidden variable models, namely binary factor analysis, topic models (latent Dirichlet allocation, correlated topic model), the restricted Boltzmann machine and the deep belief net. We shortly present these models and outline their estimation. Performance is measured by log likelihood values of these models for a holdout data set of market baskets. For each model we estimate and evaluate variants with increasing numbers of hidden variables. Binary factor analysis vastly outperforms topic models. The restricted Boltzmann machine and the deep belief net on the other hand attain a similar performance advantage over binary factor analysis. For each model we interpret the relationships between the most important hidden variables and observed category purchases. To demonstrate managerial implications we compute relative basket size increase due to promoting each category for the better performing models. Recommendations based on the restricted Boltzmann machine and the deep belief net not only have lower uncertainty due to their statistical performance, they also have more managerial appeal than those derived for binary factor analysis. The impressive performances of the restricted Boltzmann machine and the deep belief net suggest to continue research by extending these models, e.g., by including marketing variables as predictors.

Keywords: market basket analysis, factor analysis, topic models, deep learning.

Income Inequality and Price Elasticity of Market Demand: The Case of Crossing Lorenz Curves

Marat Ibragimov[1], Rustam Ibragimov[2,3], Paul Kattuman[4], Jun Ma[4]
[1]Kazan (Volga Region) Federal University, Russia,
[2]Imperial College Business School, United Kingdom, [3]Innopolis University, Russia, [4]Judge Business School, University of Cambridge, United Kingdom

This paper is concerned with the relationship between price elasticity of market demand and income inequality. Ibragimov and Ibragimov (2007) consider the market demand function for a good, aggregated from individual demand functions with price and income as arguments. Under the assumption of common preferences that are independent of income levels, they characterize the changes in price elasticity of market demand with changes in income distribution. For the limited class of changes in income distribution where increases and decreases in income inequality occur through non-intersecting Lorenz curve shifts, Ibragimov and Ibragimov (2007) show how the increase or decrease in the price elasticity of market demand depends on the Schur-convexity / Schur-concavity of the demand function.

We extend these results, establishing sufficient conditions under which price elasticity of market demand increases or decreases under general changes in income distribution, allowing Lorenz curves to intersect as they shift. Our analysis also ties in the implications of different types of tax policy changes. In practice changes in tax schedules do not lead to non-intersecting Lorenz curve shifts. We trace the price elasticity implications of a class of changes in direct tax schedules.

Keywords: Income distribution, Inequality, Downside inequality aversion, Market demand elasticity, Direct tax policy.

Optimal Bundling Strategies for Complements and Substitutes with Heavy-Tailed Valuations and Related Problems

Rustam Ibragimov[1,2], Artem Prokhorov[3,4], Johan Walden[5]
[1]Imperial College Business School, United Kingdom, [2]Innopolis University, Russia, [3]The University of Sydney Business School, Australia, [4]St. Petersburg State University, Russia, [5]Haas School of Business, the University of California at Berkeley, United States

Using majorization theory and stochastic inequalities and extending earlier works in the field, we develop a framework that allows one to model the optimal bundling problem of a multiproduct monopolist providing interrelated goods with an arbitrary degree of complementarity or substitutability. Characterizations of optimal bundling strategies are derived for the seller in the case of heavy-tailed valuations and tastes for the products. We show, in particular, that if goods provided in a Vickrey auction or any other revenue equivalent auction are substitutes and bidders' tastes for the objects are moderately heavy-tailed, then the monopolist prefers separate provision of the products. However, if the goods are complements and consumers' tastes are extremely thick-tailed, then the seller prefers providing the products on a single auction. We also present results on consumers' preferences over bundled auctions for complements and substituties in the case when their valuations exhibit heavy-tailedness. In addition, we obtain characterizations of optimal bundling strategies for a monopolist who provides complements or substitutes for profit-maximizing prices to buyers with heavy-tailed tastes. The results and approaches presented in the paper are applicable in a number of other fields in economics, finance, risk managements and related fields, including the analysis of diversification optimality under heavy-tailedness and dependence, comparisons of voting mechanisms, robust statistical and econometric methods and other areas.

Keywords: Optimal bundling, Vickrey auction, substitutes, complements, heavy-tailedness, valuations, tastes, robustness.

Some Remarks on Limit Results of the Theory of Discrete Time Branching Processes

Azam A. Imomov
State Testing Centre under the Cabinet of Ministers of the Republic of Uzbekistan, Uzbekistan

Let $\{Z_n,\ n \geq 0\}$ denote the Galton-Watson branching process having offspring probability generating function $F(s) = \sum_{j\geq 1} p_j s^j$. Assume that $p_0 > 0$ and the mean per-capita offspring number is $\sum_{j\geq 1} jp_j = 1.$, i.e. the process is critical. We also suppose that

$$F(s) = s + (1-s)^{1+\nu}\mathcal{L}\left(\frac{1}{1-s}\right) \qquad [\mathcal{R}_\nu]$$

For $0 \leq s < 1$, where $0 \leq \nu < 1$ and the function $\mathcal{L}(.)$ is slowly varying at infinity. Let the variable $\mathcal{H}$ be an extinction time of the process and $Q_n := P\{\mathcal{H} > n\}$. Slack [2] has proved that the distribution function $G_n(x) = P\{Q_n Z_n \leq x | \mathcal{H} > n\}$ converges weakly to a nonnegative limit law which has the Laplace transform in the form of $\psi(\theta) = 1 - (1+\theta^{-\nu})^{-1/\nu}$. In this report we improve the previous results provided that the function $\mathcal{L}(.)$ is slowly varying with remainder, namely

$$\frac{\mathcal{L}(\lambda s)}{\mathcal{L}(s)} = 1 + o\left(\frac{\mathcal{L}(x)}{x^\nu}\right), \qquad [\mathcal{L}_\nu]$$

for each $\lambda > 0$; see [1, p.185].

Theorem. *Let* $Z_0 = 1$. *If the conditions* $[\mathcal{R}_\nu]$ *and* $[\mathcal{L}_\nu]$*, then*

$$n^{1+1/\nu}.P\{Z_n = j\} = \frac{N(n)}{p_0}\left(1 - \frac{1+\nu}{2\nu^2}\frac{\ln n}{n}(1+o(1))\right),$$

where $N^\nu(n)\left((\nu n)^{1/\nu}/N(n)\right) \to 1$ *and*

$$\sup_{\theta>0}\left|\int_{x>0} e^{-\theta x} dG_n(x) - \psi(\theta)\right| = \frac{1+\nu}{2\nu^2}\frac{\ln n}{n}\ 1 + o(1)\ .$$

Keywords. Branching process, slowly varying function, extinction time, limit theorems.

References:

[1]. Bingham N.H., Goldie C.M. and Teugels J.L., *Regularly Variation*. University Press, Cambridge, 1987.

[2]. Slack R.S., *Further notes on branching processes with mean 1*. Wahrscheinlichkeitstheor. und Verv. Geb., 25 (1972), 31–38.

Short Term Forecasting for Electricity Demand in Egypt using Artificial Neural Networks

Mohamed A. Ismail, Alyaa R. Zahran, Eman M. Abd El-Metaal
Dept of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt

Electricity is important for any nation. It influences not only the economy, but also the political and social aspects of a nation. Forecasting electricity demand is vital for future technical improvements. Short-term electricity demand forecasts are important for controlling of the electric power system. Recently, electricity demand series has found to contain more than one seasonal pattern. Intraday and intraweek seasonal patterns are appeared in the Egyptian electricity demand time series. This study investigates using Artificial Neural Networks in accommodating these seasonality patterns for forecasting hourly electricity demand in Egypt by using seasonal lags as inputs. Different artificial neural networks with different seasonal daily and weekly lags are used. The mean absolute percentage error is used to compare forecasting power of different artificial neural networks. Results indicate the accuracy of forecasts produced by the different artificial neural networks for different time horizons.

Keywords: Electricity demand forecasting, Mean Absolute Percentage Error, Artificial Neural Networks, Double Seasonality.

Why Elections are Hard: A Game Theoretic Examination of Complex Strategic Interactions Among Multiple Political Candidates

Meredithe A. Jessup II, Darryl K. Ahner
Air Force Institute of Technology, USA

We consider a hypothetical political election in which 4 candidates compete for a nomination in a series of ballots. We show that complex interactions may preclude a Nash equilibrium as candidates may gradually adjust their platforms, within reasonable limits, between political events to appeal to a collection of distinct voter groups. We quantify candidate platforms and voter group preferences then use game theoretic principles and optimization techniques to explore the impact of platform changes within a strategic game. For the game presented here, we conjecture that no Nash equilibrium exists. Therefore, the timing of strategy changes take on a time component that better lends the problem to be solved as an extensive game.

Keywords: Election, Game Theory, Extensive Games, Primary Election.

Dependent Credit-Rating Migrations: a Heuristics for Estimating Unknown Parameters

Y. Kaniovski[1], Y. Kaniovskyi[2], G. Pflug[3]
[1]Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy, [2]Scientific Computing Research Group, Faculty of Computer Science, University of Vienna, Austria, [3]Dept of Statistics and Decision Support Systems, University of Vienna, Austria

Modeling dependent credit-rating migrations with coupling schemes, there are $\mathrm{M} \times S + 2^{\mathrm{M} \times S}$ parameters to estimate if debtors are classified into M non-default credit-classes and S industries. For a typical choice of M and S, this turns out to be a hard task for a desktop computer and a standard solver. A heuristics is suggested such that: initially a simplified problem with $\mathrm{M} \times S + 2^{\mathrm{M}}$ variables is solved, thereafter, the number of unknowns does not exceed a couple of hundreds.
For M = 2 and S = 6, two models of dependent credit-rating migrations and the respective maximum likelihood estimators are tested on Standard and Poor's (S&P's) data. Using MATLAB optimization software, exact solutions and their heuristic approximations are evaluated.
Math. Subj. Class. (2000): 90C30, 90C90

Optimal Policies for MDPs with Unknown Parameters

Michael N. Katehakis
Rutgers University, USA

We will give a brief survey of the state of the art of the area of computing optimal data driven policies for MDPs with unknown transition probabilities and or rewards. Then, we will present two simple algorithms for optimizing the average reward in an unknown irreducible MDP. The main idea of the first algorithm is to use estimates for the MDP and to choose actions by maximizing an inflation of the estimated right hand side of the average reward optimality equations. The second algorithm is based on estimating the optimal rates at which actions should be taken. For the first we show that the total expected reward obtained by this algorithm up to time n is within O(ln n) of the reward of the optimal policy, and in fact it achieves asymptotically minimal regret. Various computational challenges and simplifications are discussed.

Talk based on joint work with Wesley Cowan, PhD, Rutgers University.

Monitoring the Compliance of Countries on Emissions Abatement Levels, using Dissimilarity Indices

Ketzaki Eleni[1], Rallakis Stavros[2], Farmakis Nikolaos[1],
[1]*Dept of Mathematics, Aristotle University of Thessaloniki, Greece,*
[2]*Dept of Business Administration, University of Macedonia, Greece*

The gaseous pollutant emissions are considered as the major environmental issue. It is crucial for the climate sustainability, the countries to take collective actions on greenhouse gasses emissions abatement. These actions are taken more often through International Environmental Agreements and rarely by International Institutions. One of these agreements is the 1992's Kyoto protocol which set some targets for the emissions abatement among its members. Unfortunately, until now, little progress has been made, by the member states. In this study it is proposed a method for monitoring the implementation of the compliance of countries on emissions abatement levels using dissimilarity indices. This method will examine not only the measurement of dissimilarity, but will also contribute to the identification of the "free rider" problem, which causes the non compliance of member states.

Keywords: environmental agreements, dissimilarity indices, inequality measures, Gini index, Kyoto protocol, greenhouse gasses emissions.

Asymptotic Rate for Weak Convergence of Random Walk with a Generalized Reflecting Barrier

Tahir Khaniyev[1], Basak Gever[1], Zulfiye Hanalioglu[2]
[1]*Dept of Industrial Engineering, TOBB University of Economics and Technology, Turkey,* [2]*Dept of Actuary and Risk Management, Karabuk University, Turkey*

In this study, a random walk process $(X(t))$ with a generalized reflecting barrier is constructed mathematically and under some weak conditions, the ergodicity of the process is proved. The explicit form of the ergodic distribution is found and after standardization, it is shown that the ergodic distribution converges to the limit distribution $R(x)$:

$$Q_X(\lambda x) \equiv \lim_{t\to\infty} P\{X(t) \le \lambda x\}$$

$$\xrightarrow[\lambda\to\infty]{} R(x) = \frac{2}{\mu_2}\int_0^x \left\{\int_z^\infty \big(1 - F_+(v)\big)dv\right\} dz$$

Here, $F_+(x)$ is the distribution function of the first ladder height (χ_1^+) generated by $S_n = \sum_{i=1}^n \eta_i$ and $\mu_2 = E(\chi_1^{+2})$. Moreover, $\{\eta_i, i = 1,2,\dots\}$ random sequence represents the jumps of the process $X(t)$.
Finally, in order to evaluate asymptotic rate of the weak convergence, the following inequality is obtained, when λ is sufficiently large:

$$|Q_Y(x) - R(x)| \le \frac{2\mu_1 m_1\big(1-\pi_+(x)\big) + m_2\big(1-F_+(x)\big)}{\lambda m_1 \mu_2}$$

Here, $\pi_+(x) = \frac{1}{\mu_1}\int_0^x \big(1-F_+(t)\big)\,dt; F_+(x) \equiv P\{\chi_1^+ \le x\};$
$\mu_k = E(\chi_1^{+k}), k = 1,2;\ m_1 = E(\eta_1).$

Keywords: Asymptotic rate, Random walk, Reflecting barrier, Weak convergence.

Multivariate Risk Models and Their Applications

Yury Khokhlov, Olga Rumyantseva
Dept of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Russia

In many applications we need to use multivariate risk models which have some specific properties: their components are dependent, every component has the property of long-range-dependence, the correspondent distributions have heavy tails. In our report we consider the examples of such models. To construct such models we use the general variant of so called reduction method and subordinated processes. Next we consider some applications of these models: estimation of ruin probability in multivariate collective risk model, estimation of Tail Conditional Expectation for a portfolio components, upper and lower bounds for buffer overflow probability in teletraffic theory.

Keywords: Multivariate risk models.

Acknowledgements: *This research is supported by Russian Scientific Foundation, project 14-11-00364.*

The Mathematical Modeling of the Global Climatic Migration

Talgat R. Kilmatov
Maritime State University, Russia

The demographic challenges, population distribution, migration is closely linked. The climatic changes can produce the global migration flows. The heating of the planet will give the problems for the comfort of the human existence, efficient agriculture in some regions. There is the expectation of the environment improvement in other parts of the world.
A simple analytical mathematical model of the spatial redistribution of the population due to changes of the environment capacity is made. The

mathematical optimization method with some constraints is used (Lagrange's method). The objective function is to maximize the total gross domestic product. The constraint is the population limitation in all regions. There is possibility to change the model objective function and the constraints. The internal model parameters are the environment capacity, productivity, quantity and quality of the population.
The result of the simulation is to compare the change between the initial and final spatial of the population distribution due to changes in the natural environment capacity. The central problem of modeling is the simulation of an adequate assessment of environment capacity changes, choosing of the objective function, the population ability to the migration.

Keywords: climatic change, migration, environment capacity.

Quantile Spectral Analysis of Time Series

Tobias Kley
Dept of Statistics, London School of Economics, United Kingdom

Classical methods for the spectral analysis of time series account only for covariance-related serial dependencies. In this talk, an alternative method is presented, where, instead of covariances, differences of joint copulas and the independence copula are used to quantify serial dependencies. The Fourier transformation of these quantities is considered and used to define new spectral quantities: the *quantile (cross) spectral density kernel* and the *quantile coherency kernel*. They allow to separate marginal and serial aspects of a time series and intrinsically provide more information about the conditional distribution than the classical location-scale model. Thus, the quantile spectral density and quantile coherency kernel are more informative than the spectral density obtained from the autocovariances. For an observed time series the quantile spectral density and coherency kernel are estimated. The asymptotic properties, including the order of the bias and process convergence, of the $\ell^\infty([0,1]^2)$-valued estimator are established. The results are applicable without restrictive distributional assumptions such as the existence of finite moments and only a weak form of mixing, such as α-mixing, is required.

Keywords: Copula, Periodogram, Quantile coherency, Ranks.

This presentation is based on joint work with S. Volgushev, H. Dette, M. Hallin (Bernoulli 22(3), 2016), and J. Baruník (arXiv:1510.06946).

Asymptotic Confidence Regions of Parameters of the Skew Normal Distribution

Tõnu Kollo, Meelis Käärik, Anne Selart
Institute of Mathematics and Statistics, University of Tartu, Estonia

The asymptotic normality is established for the estimators of the shape vector and the scale matrix of the two-parameter multivariate skew normal distribution for two parameterizations. The used point estimators have been found by the method of moments. Also, an analytic expression and an asymptotic normal law are derived for the estimator of the skewness vector of the skew normal distribution. The expressions of the moments in matrix representation are used in derivation. Convergence to the asymptotic distributions is examined both computationally and in a simulation experiment.

Keywords: asymptotic normality, multivariate cumulants, multivariate moments, multivariate skewness, skew normal distribution.

Operating Environment Threats Influence on Critical Infrastructure Safety – The Numerical Approach

Krzysztof Kołowrocki, Ewa Kuligowska, Joanna Soszyńska-Budny
Gdynia Maritime University, Poland

The material given in this paper delivers the procedure for numerical approach that allows finding the main practically important safety characteristics of the critical infrastructures at the variable operation conditions including operating environment threats. The obtained results are applied to the safety evaluation of the port oil piping transportation system. It is assumed that the conditional safety functions are different at various operation states and have the exponential forms. Using the procedure and the program written in Mathematica, the considered port oil piping transportation system main characteristics including: the conditional and the unconditional expected values and standard deviations of the system lifetimes, the unconditional safety function and the risk function are determined.

Keywords: safety, operating environment threat, port oil piping transportation system.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Port Oil Transport Critical Infrastructure Safety Approximate Evaluation

Krzysztof Kołowrocki, Joanna Soszyńska-Budny
Gdynia Maritime University, Poland

The method based on the multistate approach to critical infrastructure safety modelling is proposed and practically useful critical infrastructure safety indicators are created. The proposed method is applied to the safety analysis of the port oil piping transportation system. Safety indicators of this critical infrastructure are approximately evaluated on the basis of data coming from experts.

Keywords: critical infrastructure, safety, safety indicator, prediction, port oil transport.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Safety Prediction of Critical Infrastructure Impacted by Climate-Weather Change Process

Krzysztof Kołowrocki, Joanna Soszyńska-Budny
Gdynia Maritime University, Poland

The paper is devoted to the climate influence on the safety of a critical infrastructure defined as a complex system in its operating environment that in the case of its degradation have significant destructive influence on the health, safety and security, economics and social conditions of large human communities and territory areas. The method based on the joint model linking a multistate approach to critical infrastructure safety with a semi-Markov modelling of the climate-weather change process at the critical infrastructure operation area is proposed to the safety analysis and prediction of critical infrastructures impacted by the climate hazards.

Keywords: critical infrastructure, climate change, safety, climate impact, safety indicator, prediction, port oil transport.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Simplified Approach to Safety Prediction of Port Oil Transport Critical Infrastructure Related to Operation Process

Krzysztof Kołowrocki, Joanna Soszyńska-Budny
Gdynia Maritime University, Poland

The method based on the joint model linking a multistate approach to critical infrastructure safety with a semi-Markov modelling of the critical infrastructure operation process is proposed to the safety prediction of critical infrastructures changing in time their structure and their components safety. The proposed method is applied to the safety indicators approximate evaluation of the port oil transport critical infrastructure changing its safety structure and its components safety parameters at variable operation conditions.

Keywords: critical infrastructure, operation, safety, operation influence, safety indicator, prediction, port oil transport.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Analysis of Dependencies between Growth Rates of GDP of V4 Countries Using 4-dimensional Vine Copulas

Jozef Komorník[1], Magda Komorníková[2], Tomáš Bacigál[2]
[1]*Faculty of Management, Comenius University, Slovakia,*
[2]*Faculty of Civil Engineering, Slovak University of Technology, Slovakia*

We analyzed seasonally adjusted quarterly data (provided by EUROSTAT) for the V4 countries Poland (Pl), Czech Republic (Cz), Hungary (Hu) and Slovakia (Sk) for the period 2002/Q1 - 2016/ Q2. We applied ARIMA - GARCH filters to logarithms of the above countries data. The obtained residuals have pairwise correlation coefficients in the interval (0.35, 0.7) (the maximal values were achieved for the couples (Pl, Cz) and (Pl, Hu)). Subsequently, we applied to those residuals (country specific) linear transformations in order to map them in the unit interval. The results served as inputs to calculation of 4-dimensional vine copulas. The best 4-dimensional vine copula (based on the BIC information criterion) was:

Tree 1:

Pl - Hu: *Survival Joe* (par = 15.94, Kendall's tau = 0.88)
Sk - Cz: *Frank* (par = 12.09, Kendall's tau = 0.71)
Sk - Pl: *Frank* (par = 9.7, Kendall's tau = 0.66)

Tree 2:

(Sk, Pl) - (Hu, Pl): *Joe* (par = 16.79, Kendall's tau = 0.89)
(Pl, Sk) - (Cz, Sk): *Survival Joe* (par = 15.94, Kendall's tau = 0.88)
Tree 3:
(Cz, Sk,Pl)-(Hu, Sk,Pl): *t* (par_1 = 0.12, par_2 = 11.02, Kendall's tau = 0.08)
Since vine copulas are constructed as lego using bivariate copulas as construction blocks picked by default for stronger correlated random variable pairs, the estimated model structure can be visualized and interpreted quite easily (yet with care). The lowest tree reveal stronger link between Slovak and Czech, Slovak and Polish, and Polish and Hungarian GDP, but only the later indicates tail dependence related to smallest values.
The same holds for distribution of (Pl, Cz) random vector conditional on Sk visible in the middle tree. Parallely, relation between Sk and Hu conditional on Hu displays mirrored behavior (with upper tail dependence). The top level tree represents weak or no dependence when conditioning on Sk and Pl.

Keywords: GDP, Correlation, ARIMA-GARCH filter, Vine copula.

Acknowledgements: *The support of the grants APVV-0013-14 and VEGA 1/0420/15 is kindly announced.*

Exact Lower Bounds for the Agnostic Probably-Approximately-Correct (PAC) Machine Learning Model

Aryeh Kontorovich[1], Iosif Pinelis[2]
[1]*Dept of Computer Science, Ben-Gurion University, Israel,*
[2]*Dept of Mathematical Sciences, Michigan Technological University, U.S.A.*

We provide an exact asymptotic lower bound on the minimax expected excess risk (EER) in the agnostic probably-approximately-correct (PAC) machine learning classification model. This bound is of the simple form $c_\infty/\sqrt{\nu}$ as $\nu \to \infty$, where $c_\infty = 0.16997...$ is a universal constant, $\nu = m/d$, m is the size of the training sample, and d is the Vapnik–Chervonenkis dimension of the hypothesis class. In the case when randomization of learning algorithms is allowed, we also provide an exact non-asymptotic lower bound on the minimax EER and identify minimax learning algorithms as certain maximally symmetric and minimally randomized "voting" procedures. It is shown that the differences between these asymptotic and non-asymptotic bounds, as well as the differences between these two bounds and the maximum EER of any learning algorithms that minimize the empirical risk, are asymptotically negligible, and all these differences are due to ties in the mentioned "voting" procedures. A few easy to compute non-asymptotic lower bounds on the minimax EER are also obtained, which are shown to be close to the

exact asymptotic lower bound $c_\infty/\sqrt{\nu}$ even for rather small values of the ratio $v = m/d$. As an application of these results, we substantially improve existing lower bounds on the tail probability of the excess risk. Among the tools used are Bayes estimation and apparently new identities and inequalities for binomial distributions.

Mathematical Aspects of the Nuclear Glory Phenomenon: from Backward Focusing to Chebyshev Polynomials

Vladimir B. Kopeliovich[1;2]
[1]*Institute for Nuclear Research of RAS, Russia,*
[2]*Moscow Institute of Physics and Technology (MIPT), Russia*

The angular dependence of the cumulative particles production off nuclei near the kinematical boundary for multistep process is defined by characteristic polynomials in angular variables, describing spatial momenta of the particles in intermediate states [1, 2]. Physical argumentation, exploring the small phase space method, leads to the appearance of equations for these polynomials in $\cos(\theta/\mathrm{N})$, where θ is the polar angle defining the momentum of final (cumulative) particle, the integer N being the multiplicity of the process (the number of interactions). It is shown explicitly how these equations aappear, the recurrent relations between polynomials with different N are obtained. Factorization properties of characteristic polynomials found in our previous work [3] are extended, and their connection with known in mathematics Chebyshev polynomials of 2-d kind [4, 5] is established. As a result, differential cross section of the cumulative particle production has characteristic behavior $d\sigma \sim 1/\sqrt{\pi - \theta}$ near the strictly backward direction ($\theta = \pi$, the backward focusing effect). Such behaviour takes place for any multiplicity of the interaction, beginning with $n = 3$, elastic or inelastic (with resonance excitations in intermediate states) and can be called the nuclear glory phenomenon, or 'Buddha's light' of cumulative particles.

References:

[1]. V.B. Kopeliovich. Sov.J.Nucl.Phys. 26 (1977) 87 [Yad.Fiz. 26 (1977) 168]
[2]. V.B. Kopeliovich. Phys. Rept. 139 (1986) 51
[3]. V.B. Kopeliovich, G.K. Matushko and I.K. Potashnikova, J.Phys. G41 (2014) 125107; Vladimir Kopeliovich and Galina Matushko. 'Buddha's light' of cumulative particles. Nuclear Glory Phenomenon. Scholars Press, Saarbruecken, Germany. 2015. ISBN 978-3-639-76496-3; V. Kopeliovich, arXiv:1610.09776 [nucl-th]
[4]. Chebyshev polynomials. https://en.wikipedia.org/wiki/Chebyshevpolynomials
[5]. Mnogochleny Chebysheva. https://ru.wikipedia.org/wiki/Mnogochleny-Chebysheva

Rescaling of Quantized Skyrmions: from Nucleon to Baryons with Heavy Flavor

Vladimir B. Kopeliovich[1;2], Irina K. Potashnikovady[3]
[1]Institute for Nuclear Research of RAS, Russia, [2]Moscow Institute of Physics and Technology (MIPT), Russia, [3]Departamento de Física, Universidad Técnica Federico Santa María and Centro Científco-Tecnológico de Valparaíso, Chile

The role of rescaling (expansion or squeezing) of quantized skyrmions [1, 2] is studied for the spectrum of baryons beginning with nucleon and Δ(1232), and with flavors strangeness, charm or beauty. The expansion of skyrmions due to the centrifugal forces has influence on the masses of baryons without flavor (N and especially Δ) and demands certain revision of the fit of the model proposed in [3, 4]. The rescaling of skyrmions has smaller influence on the spectrum of strange baryons, it is more important for the case of charm, and is crucial for baryons with beauty quantum number, where strong squeezing takes place [5]. Two competing tendencies are clearly observed: expansion of skyrmions when isospin (or spin) increases, and squeezing with increasing mass of the flavor. For the case of beauty baryon Λ_b satisfactory agreement with data can be reached for the value $r_b = F_B/F_\pi \simeq 2.6$, for the case of Σ_b there should be $r_b \sim 2$, so for the beauty flavor the method seems to be not quite satisfactory because of certain intrinsic discrepances. Some pentaquark states with hidden strangeness, charm or beauty are considered as well [5].

References:

[1]. T.H.R. Skyrme. A Nonlinear field theory. Proc.R.Soc. London, A260 (1961) 127; T.H.R. Skyrme. A Unified Field Theory of Mesons and Baryons. Nucl.Phys. 31 (1962) 556

[2]. E. Witten. Global Aspects of Current Algebra. Nucl.Phys. B223 (1983) 422; Current Algebra, Baryons, and Quark Confinement. ibid B223 (1983) 433

[3]. G. Adkins, C. Nappi, and E. Witten. Static Properties of Nucleons in the Skyrme Model. Nucl.Phys. B228 (1983) 552

[4]. G. Adkins and C. Nappi. The Skyrme Model with Pion Masses. Nucl.Phys. B233 (1984) 109

[5]. Vladimir B. Kopeliovich and Irina K. Potashnikova. Rescaling of quantized skyrmions: from nucleon to baryons with heavy avor e-Print: arXiv:1703.04373 [hep-th]

Monte-Carlo Accuracy Evaluation of Pintograph-based Laser Cutting Machine

Samuel Kosolapov
ORT Braude Academic College of Engineering, Israel

Harmonograph is a mechanical apparatus that employs pendulums to create a geometric image. Pintograph is a lateral (2D) implementation of the Harmonograph: it uses a number of rods to move a pen (or some tool) relative to a flat drawing surface. Known from 19th century, Pintograph became popular our days because of its simple mechanical implementations which uses inexpensive servo motors controlled by an inexpensive microcontroller. Despite mechanical design is simple, mathematical model of the real-life Pintograph contains a big number of mechanical and electronic parameters. Hence, in order to evaluate accuracy of the Pintograph, Monte-Carlo software simulator was created. Relevant math equations were created and solved by using MAPLE symbolic software. Simulator takes into account length of a rods, joint' tolerances and servo motors accuracy and resolution. Simulator operation results in the drawing zone sha! pe and in the accuracy map in the drawing zone. Simulator runs reveal that for the "5-joints" design with unit length 100 mm xy accuracy of 0.3 mm can be achieved in the center of the drawing zone, which may be good enough for an inexpensive laser cutting DIY machine. In case better accuracy is required, required sizes and tolerances of the Pintograph elements can be evaluated.

Keywords: Monte-Carlo simulator, Harmonograph, Pintograph, MAPLE, Laser cutting.

Compound Distributions Associated with Order Statistics

Markos V. Koutras, Vasileios M. Koutras
Department of Statistics and Insurance Science, University of Piraeus, Greece

In the present work we study a compound order statistics distribution generated by considering order statistics of a random number of random variables. Let N be a non-negative integer valued random variable and $X_1, X_2, \ldots$ a sequence of independent and identically distributed random variables, independent of N. Our interest focuses on the distribution of the *r*-th order statistic $X_{r:N}$ of a random sample $X_1, X_2, \ldots, X_N$ whose length N is a random variable.

Some results pertaining to the exact distribution of $X_{r:N}$ are presented and the properties of the distribution are studied in some detail. Besides the general setup, several interesting outcomes are developed when *N* belongs to wide classes of discrete distributions such as the family of discrete phase-type distributions or other parametric families of discrete distributions.

Finally, we illustrate how the stochastic model under study can be exploited for modeling problems arising in several areas of applied

science such as actuarial science, biostatistics, educational psychology, financial risk management, quality control and reliability, etc.

Keywords: Compound distributions; random order statistics; phase-type distributions; waiting times; risk management; biostatistics; statistical quality control.

Acknowledgment: *Work funded by National Matching Funds 2014-2016 of the Greek Government, and more specifically by the General Secretariat for Research and Technology (GSRT), related to EU project "IS MPH: Inference for a Semi-Markov Process" (GA No 329128).*

Financial Risk Management Modeling via Random Order Statistics

V. M. Koutras, M. V. Koutras
Department of Statistics and Insurance Science, University of Piraeus, Greece

Several applications related to risk management problems and especially financial risk analysis can be studied by the aid of stochastic models involving the minimum, maximum or more generally a specific order statistic of a random sample of random size. For example in the credit risk area a mortgage portfolio the number of loans that will default is a random variable, while, at the same time, the loss given default for each of them is a random variable as well. Under this scenario, a financial institution might be interested at the stochastic behavior of the minimum, maximum, median loss in the portfolio, or more generally the *r*-th largest loss in it. The motivation of the model studied in the present work stems from the aforementioned application and similar in nature applications arising in several areas of applied science such as bank supervision, bank capitalization monitoring, actuarial science, quality control monitoring etc.

Let T denote the waiting time for the first occurrence of a specific outcome in a sequence of multinomial trials and $X_1, X_2, \ldots$ a sequence of independent and identically distributed random variables, which are independent of T. In the present article we develop some results on the distribution of the compound random variables $\min\{X_1, X_2, \ldots, X_T\}$, $\max\{X_1, X_2, \ldots, X_T\}$ and illustrate how these results can be profitably used to study models pertaining to financial risk management problems.

Keywords: Waiting times; compound distributions; pattern statistics; bank supervision; risk management.

Acknowledgment: *Work funded by National Matching Funds 2014-2016 of the Greek Government, and more specifically by the General Secretariat for Research and Technology (GSRT), related to EU project "IS MPH: Inference for a Semi-Markov Process" (GA No 329128).*

The Use of Deviance Plots for Non-Nested Model Selection in Diverse Models

Pieter M. Kroonenberg
Dept of Education and Child Studies, Leiden University, The Netherlands
& The Three-Mode Company, The Netherlands

Mallows defined the Cp statistic with an associated Cp plot to be used in model selection in regression analysis. The deviance plot is an extension of this idea, where the loss, expressed in residual sum of square, or the Chi-squared statistic is graphed against the degrees of- freedom, thus allowing for comparing deviance/df ratios between models. It is shown that both the RMSEA (Root Mean Squared Error of Approximation) and the AIC (Akaike's Information Criterion) are lines in the Chi-squared deviance plot so that these criteria can be used together. Moreover work has been done for automatic selection by finding the point where the convex hull of the optimal solution shows the largest curvature. Brief examples for different methods will be provided.

Keywords: Model comparison, regression, loglinear analysis, structural equation modelling, three-mode analysis.

Climate-Weather Change Process Realizations Uniformity Testing for Maritime Ferry Operating Area

Ewa Kuligowska, Mateusz Torbicki
Gdynia Maritime University, Poland

The paper is concerned with a method for statistical data uniformity testing applied to the empirical conditional sojourn times' realizations of the climate-weather change process at the particular climate-weather states for maritime ferry operating at Baltic Sea open waters. The empirical data sets during Februaries of the years 1988-1993 at the four different measurement points are collected. Assuming that the statistical data sets are separate, the verification of the non-parametric hypotheses on the basis of Kolmogorov-Smirnov test and Wald-Wolfowitz runs test is prepared. In the case when the null hypothesis about the uniformity data sets for two measurement points is not rejected, corresponding to each other statistical data sets are joined. Next, the one joined climate-weather change process for those two measurement points is considered. After using this approach, the climate-weather change process at the fixed area is better described.

Keywords: climate-weather change process, uniformity testing, maritime ferry.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Climate-Weather Change Process Realizations Uniformity Testing for Port Oil Piping Transportation System Operating Area

Ewa Kuligowska, Mateusz Torbicki
Gdynia Maritime University, Poland

The paper is concerned with a method for statistical data uniformity testing applied to the empirical conditional sojourn times' realizations of the climate-weather change process at the particular climate-weather states for port oil piping transportation system operating under Baltic Sea waters area. The empirical data sets during Februaries of the years 1988-1993 at the four different measurement points are collected. Assuming that the statistical data sets are separate, the verification of the non-parametric hypotheses on the basis of Kolmogorov-Smirnov test and Wald-Wolfowitz runs test is prepared. In the case when the null hypothesis about the uniformity data sets for two measurement points is not rejected, corresponding to each other statistical data sets are joined. Next, the one joined climate-weather change process for those two measurement points is considered. After using this approach, the climate-weather change process at the fixed area is better described.

Keywords: climate-weather change process, uniformity testing, port oil piping transportation system.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Identification and Prediction of Climate-Weather Change Process for Port Oil Piping Transportation System and Maritime Ferry Operation Areas after Successful Uniformity Testing

Ewa Kuligowska, Mateusz Torbicki
Gdynia Maritime University, Poland

The paper is concerned with a practical application of unknown joined climate-weather change process' parameters identification after successful uniformity testing of climate-weather data sets coming from different measurement points. The general probability model of a

climate-weather change process is introduced. The identification of the port oil piping transportation system climate-weather change process and the maritime ferry climate-weather change process is performed for the uniformly tested joined statistical data sets of the corresponding climate-weather change processes' sojourn times realisations. Finally, those identified climate-weather change processes are applied to the climate-weather characteristics prediction.

Keywords: climate-weather change process, port oil piping transportation system, maritime ferry.

Acknowledgments: *Paper in the scope of EU-CIRCLE project funded from the European Union's Horizon 2020 Research and Innovation Programme under grant Agreement No 653824.*

Effect of Statin Prescription and on Population Life Expectancy

Elena Kulinskaya, Lisanne Gitsels, Ilyas Bakbergenuly
School of Computing Sciences, University of East Anglia, United Kingdom

Longevity and morbidity risks are of essential importance to the actuarial community. We believe that to be able to establish the drivers of their changes, and to predict how they may change over time and how this would affect life expectancy, researchers need to engage in statistical modelling of mortality experience using large scale population-based individual level data collected over the long term. To determine the main factors affecting longevity and dynamics of their changes, we are using the subset of the THIN primary care database comprising 3.4 million patients born before 1960. We provide a case-study on longevity improvement due to the widening of the prescription of statins, the only known longevity-improving drug in general use. Joint modelling of statin prescription and survival experience allows to develop a plausible scenario of temporal changes in longevity due to statins.

Keywords: Joint modelling, longitudinal measurements, survival analysis, primary care data.

On multiple Step Stress Model under Order Restriction

Debasis Kundu
Dept of Mathematics and Statistics, Indian Institute of Technology Kanpur, India

In this article we consider multiple step stress model based on the cumulative exposure model assumption. Here we assume that for a given stress level, lifetime of the experimental units follow exponential

distribution and the expected lifetime decreases as the stress level increases. We mainly focus on the order restricted inference of the unknown parameters of the lifetime distributions. First we consider the order restricted maximum likelihood estimators of model parameters and the associated confidence intervals. The Bayes estimates and associated credible intervals under square error loss function are also provided in this paper. Due to the absence of explicit form of Bayes estimates, we propose to use importance sampling technique to compute the Bayes estimates. We provide an extensive simulation study in case of three stress levels mainly to see the performances of the proposed methods. Finally we provide the analyses of two data sets for illustrative purposes.

This is a joint work with Mr. Debashish Samanta and Dr. Ayon Ganguly.

Residuals for the Modelling of Covariate Effects in Accelerated Life Models

Jan Terje Kvaløy[1], Bo Henry Lindqvist[2], Stein Aaserud[3]
[1]Dept of Mathematics and Natural Sciences, University of Stavanger, Norway, [2]Dept of Mathematical Sciences, Norwegian University of Science and Technology, Norway, [3]Cancer Registry of Norway, Norway

In parametric accelerated life regression models one issue of concern is proper modelling of the functional form of the covariates. We will discuss residual plots to reveal and check the functional form of the covariates in such models. Two approaches for handling censoring will be discussed; adjusting censored residuals by adding a residual time, and using nonparametric exponential regression on unadjusted censored residuals. It will in particular be demonstrated how residuals can be used to infer the correct functional form for misspecified covariates. Applications to real and simulated data will be shown.

Keywords: Standardized residuals, Cox-Snell residuals, exponential regression, misspecified model

A Stochastic Single Vehicle Routing Problem with a Predefined Sequence of Customers and Delivery of Two Similar Products

Epaminondas G. Kyriakidis[1], Theodosis D. Dimitrakos[2], Constatinos Karamatsoukis[3]

[1]Dept of Statistics, Athens University of Economics and Business, Greece,
[2]Dept of Mathematics, University of the Aegean, Greece,
[3]Dept of Military Sciences, Hellenic Military Academy, Greece

We develop and analyze a mathematical model for a specific vehicle routing problem in which a vehicle starts its route from a depot loaded with items of two similar but not identical products. We name these products, product 1 and product 2. The vehicle must deliver the products to *N* customers according to a predefined sequence. This means that first customer 1 must be serviced, then customer 2 must be serviced, then customer 3 must be serviced and so on. The vehicle has finite capacity and after servicing all customers it returns to the depot. It is assumed that each customer prefers either product 1 or product 2 with known probabilities. The actual preference of each customer becomes known when the vehicle visits the customer. It is also assumed that the quantity that each customer demands is a random variable with known distribution. The actual demand is revealed upon the vehicle's arrival at customer's site. The demand of each customer cannot exceed the vehicle capacity and the vehicle is allowed during its route to return to the depot to restock with quantities of both products. The travel costs between consecutive customers and the travel costs between the customers and the depot are known. If there is shortage for the desired product it is permitted to deliver the other product at a reduced price. The objective is to find the optimal routing strategy, i.e. the routing strategy that minimizes the expected total cost among all possible strategies. It is possible to find the optimal routing strategy using a suitable stochastic dynamic programming algorithm. It is also possible to prove that the optimal routing strategy has a specific threshold-type structure.

Keywords: Vehicle routing problem, Similar products, Stochastic dynamic programming.

Some Results on the Detection of CpG Islands via Hidden Semi Markov Modelling

Zacharias Kyritsis[1], Aleka Papadopoulou[1], Sally McClean[2]
[1]*Department of Mathematics, Aristotle University of Thessaloniki, Greece,*
[2]*School of Computing and Information Engineering, University of Ulster, Northern Ireland, UK*

In the present paper, we study DNA sequences in order to detect areas with CpG islands. In terms of hidden semi-Markov modelling we consider a state space for the hidden semi Markov process which categorizes areas of the DNA sequence according to the presence or absence of CpG islands. Also, the holding time distributions for presence or the absence states describe the length of the areas with or without CpG islands. The observation sequence is a sequence of letters in the 4-letter DNA alphabet {A, C, G, T}. A method is applied according to which we can detect areas of CpG islands in DNA sequences via hidden semi Markov modelling. Last, DNA data are applied to illustrate the above modelling.

Keywords: DNA sequences, CpG islands, hidden semi-Markov model.

Planning and Management of Transport Resources Constraints

Abdelkader Lamamri[1], Imene Mehamdia[2]
[1]*University of Djilali bounaama, Algeria,* [2]*University Kasdi Merbah, Algeria*

We are interested in problems from combinatorial optimization, more precisely, the problem of construction crew rotations with resource constraints. The problem is to cover cost of all flights of the company. Given the large size of the problems encountered in industry, 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. We present a new approach to improve the acceleration of the method of column generation problem for the construction crew rotations, it is projected in each arc, the resources on a vector of dimension less using a Lagrangean relaxation algorithm to determine the coefficients of the projection arc combined with an algorithm for re-optimization, well generates a sub-set of complementary solutions to the master problem.

Keywords: Combinatorial optimization, mathematical programming, planning and management.

Resolving a Dynamic Winner Determination Problem (WDP) by Dynamic Programming

Asli Larbi, Aider Méziane
Bejaia Université, Algeria

In this paper, we propose a mathematical formulation model for the dynamic winner determination problem (WDP), in combinatorial auctions. This model is developed for the purpose of implementing the web, the model reflects the English auction mechanism. We have developed an optimal resolution algorithm, which is based on the temporal analysis of the bid status in order to avoid to the maximum, to the combinatorial resolution. We use the dynamic programming if need be, According to state of the bid. The algorithm gives at every time of the auction period, a temporary list of winners, Articles sold and the temporary gain. A numerical experiment using this algorithm on simulated data gives rise to satisfactory results.

Keywords: English combinatorial auction, dynamic programming, temporary winners list, final winner, winner determination problem, WDP.

Perturbation of the Malliavin Calculus of Bismut Type of Large Order

Rémi Léandre
Laboratoire de Mathématiques, Université de Franche-Comté, France

In the qualitative theory of an elliptic operator, only the main term (which is given by its principal symbol) plays roughly speaking a role. We show that this statement is true for the Malliavin Calculus of Bismut type of big order on a Lie group.

Goodness-of-Fit Tests based on Entropy. Application to DNA Replication.

Justine Lequesne[1], Valérie Girardin[2], Philippe Regnault[3]
[1]Service de Recherche Clinique, Centre Henri Becquerel, France, [2]Laboratoire de Mathématiques Nicolas Oresme, UMR 6139, Université de Caen Normandie, France, [3]Laboratoire de Mathématiques de Reims, EA 4535, Université de Reims Champagne-Ardenne, France

Goodness-of-fit tests based on Shannon entropy and Kullback-Leibler divergence, that are basic concepts of information theory, are widely used in the literature.

These tests are known to have good power properties and to lead to straightforward computation for testing a large family of distributions. Mathematical justification of entropy and divergence based tests will be detailed in order to show that they provide a unique procedure for testing any parametric composite null hypothesis distribution of maximum entropy.
In addition, we have developed a package called *maxentgoftest* for the statistical software R. It provides an easy implementation of these goodness-of-fit tests for numerous families of maximum entropy distributions, including, *e.g.*, Pareto, Fisher, Weibull distributions.
The methodology and computer procedures wil be applied to a real dataset of a DNA replication program. The objective is to validate an experimental protocol to detect chicken cell lines for which the spatio-temporal program of DNA replication is not correctly executed. To this end, we propose a two-step approach through entropy-based tests, leading first to retain a Fisher distribution with non-integer parameters and then to validate the experimental protocol.

Keywords: DNA replication, goodness-of-fit tests, Kullback-Leibler divergence, R-package, Shannon entropy.

Heavy-Tailed Fractional Pearson Diffusions

N. Leonenko
Cardiff University, United Kingdom

Heavy-tailed fractional Pearson diffusions are a class of sub-diffusions with marginal heavy-tailed Pearson distributions: reciprocal gamma, Fisher-Snedecor and Student distributions. They are governed by the time-fractional diffusion equations with polynomial coefficients depending on the parameters of the corresponding Pearson distribution. We present the spectral representation of transition densities of fractional Fisher-Snedecor and reciprocal - gamma diffusions, which depend heavily on the structure of the spectrum of the infinitesimal generator of the corresponding non-fractional Pearson diffusion. Also, we present the strong solutions of the Cauchy problems associated with heavy-tailed fractional Pearson diffusions and the correlation structure of these diffusions.

Keywords: Fractional diffusion, Pearson diffusion, Heavy-tailed distribution.

References:

[1]. Leonenko, N.N., Meerschaert, M.M and Sikorskii, A. (2013) Fractional Pearson diffusions, Journal of Mathematical Analysis and Applications, vol. 403, 532-546
[2]. Leonenko, N.N., Papic, I., Sikorskii, A. and Suvak, N. (2017) Heavy-tailed fractional Pearson diffusions, Stochastic Processes and their Applications, in press

This is a joint work with I. Papic (University of Osijek, Croatia), N. Suvak (University of Osijek, Croatia) and Alla Sikorskii (Michigan State University and Arizona University, USA).

Using Child, Adult, and Old-age Mortality to Establish a Developing Countries Mortality Database (DMD)

Nan Li, Hong Mi, Patrick Gerland
School of Public Affairs/Institute for Population Studies and Development Studies, Zhejiang University, P. R. China

Life-table databases have been established for developed countries and effectively used for various purposes. For developing countries of which the deaths counted 78% that of the world in 2010-2015, however, reliable life tables can hardly be found. Indirect estimates of life tablesusing empirical data on child and adult mortality are available for developing countries. But more than half of all deaths already occurred at age 60 and higher in developing countries in 2010-2015, which leads to the irony that worldwide the number of deaths at old-ages is the biggest, and also the least reliable. This reality indicates that improving the estimates of old-age mortality for individual developing countries is not enough, and that establishing a life-table database for all developing countries, which utilizes the improved estimations of old-age mortality, is necessary.To fulfill this task,we introduce two methods: (1) the Census Method that uses populations enumerated from census to estimate old-age mortality, and (2) the three-input model life table thatutilizes child, adult, and old-age mortalityto calculate life tables. Compared to using only child and adult mortality, applying the two methods to the data from the Human Mortality Database after 1950, the errors of fitting old-age mortality are reduced for more than 70% of all the countries. For the three non-European-origin populations in the Human Mortality Database the errors are reduced by 17% for Chile, 48% for Japan, and 17% for Taiwan, which is more relevant for developing countries. These results indicate that the methodology is adequate and empirical data are available to establish a mortality database for developing countries.

Keywords: Life table, Database, Developing countries.

Insolvency as Opportunity: A Marketing Perspective on Time-Dependent Credit Risk

Caterina Liberati[1], Furio Camillo[2]
[1]Dept of Economics, Management and Statistics, Università di Milano-Bicocca Milan, Italy, [2]Dept of Statistical Science, Università di Bologna, Italy

The objective of quantitative credit scoring is to develop accurate models that can distinguish between good and bad applicants (Baesens et al, 2003). Statistical research has focused on delivering new classification methodologies based on neural networks or support vector machines that provide better predictions respect to standard classifiers but poor interpretation. A step forward in such task, it is represented by linear reconstruction of kernel discriminant proposed by Liberati et al (2015) which combines the effective classification results, due to application of no-linear functions, with an easy interpretability of the data.
In reality, insolvency, is not always a negative occurrence but the opportunity to generate more profit for a financial institution. Indeed, it can be considered a marketing leverage to reinforce customers' loyalty. In this paper, we approach credit risk modeling into this perspective. We compare the classification solution that focuses on prediction performance with an explanatory regression as Tobit that models both default probability and duration. Results will be illustrated in a double perspective (risk management and customers' segmentation), in order to identify which covariates affect the insolvent behaviour the most.

Keywords: Credit Scoring, Tobit Model, Kernel Discriminant, Time-dependent Risk.

References:

[1]. Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J (2003) Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54:627–635
[2]. Liberati C, Camillo F, Saporta G (2015) Advances in credit scoring: combining performance and interpretation in kernel discriminant analysis. *Advances in Data Analysis and Classification* (in press)
[3]. Takeshi A (1984). Tobit models: A survey. *Journal of Econometrics*. 24: 3–61.

Diffusion Approximation of a Loss Queueing System

Nikolaos Limnios
*Laboratory of Applied Mathematics (LMAC),
Université de Technologie de Compiègne, Sorbonne Universités, France*

A queueing loss system with N independent sources, without buffer, and M servers, is considered here (N>M). Arrivals and service times are Poisson and exponential distributed respectively. We present averaging

and diffusion approximation results as the number of sources and service facilities becomes together large.

Keywords: Loss queueing system, Markov processes.

A Semimartingale Characterization of Semi-Markov Processes and Branching Processes with Transport of Particles

Nikolaos Limnios[1], Elena Yarovaya[2]
[1]*Laboratory of Applied Mathematics, Université de Technologie de Compiègne, Sorbonne Universités, France,*
[2]*Dept of Probability Theory, Lomonosov Moscow State University, Russia*

Involving of a non-Markovian processes into the models of dynamical systems substantially widens the area of applications of such models, particularly in reliability theory. Among them the semi-Markov processes are the simplest generalization of Markov processes. One of the main problems in the application of semi-Markov processes is to obtain its compensator and quadratic characteristics and a solution of the equation for the marginal probabilities. To achieve this goal the method based on a jump process representation of a semi-Markov process is suggested. For applications in reliability theory it is important to reduce dynamical systems with denumerable number of states to systems with a finite number of states. We demonstrate that in such procedure Markovian property of initial dynamical systems with denumerable number of states may be lost. As an example, we consider branching processes with transport of particles called branching random walks. Non-Markovian models are constructed based on branching random walks on multidimensional lattices. The object of investigation in branching random walks is the number of particles in every lattice point. By aggregation of some lattice point areas it is possible to consider a finite set of system states instead of a denumerable set of system states, but, as a result, the Markovian property of initial branching random walks is lost. General methods are proposed to study non-Markovian models for branching random walks.

Keywords: Semi-Markov Processes, Non-Markovian Processes, Branching Random Walks, Limit Theorems.

Acknowledgements: *The research of the author[2] is partly supported by Russian Science Foundation project no. 14-21-00162.*

Majorization and Stochastic Orders in Secure Communications

Pin-Hsun Lin[1], Holger Boche[2], Eduard Jorswieck[1]
[1]TU Dresden, Germany, [2]TU München, Germany

Recently, the secrecy capacities of fading wiretap channels are derived for different types of channel state information and different system models. The case with fast fading and imperfect information at the transmitter is important for practical secure data transmission. We investigate the relation between different stochastic orders and the degradedness of a fast fading wiretap channel with statistical channel state information at the transmitter. Based on Majorization of the channel parameters, we derive sufficient conditions to identify the ergodic secrecy capacities for single- and multiple antenna wiretap systems.

References:

[1]. P.-H. Lin and E. Jorswieck, "Stochastic Orders, Alignments, and Ergodic Secrecy Capacity" in Information Theoretic Security and Privacy of Information Systems, Cambridge University Press, will appear June 2017.

[2]. P.-H. Lin, E. Jorswieck, "On the Fast Fading Gaussian Wiretap Channel with Statistical Channel State Information at the Transmitter", IEEE Trans. on Information Forensics and Security, vol. 11, no. 1, Jan. 2016.

[3]. E. Jorswieck and H. Boche, "Majorization and Matrix Monotone Functions in Wireless Communications", Foundations and Trends in Communications and Information Theory, vol. 3, no. 6, July 2007, pp. 553 - 701.

On Multi-Channel Stochastic Networks with Time-Dependent Input Flows

Hanna Livinska, Eugene Lebedev
Applied Statistics Dept,
Taras Shevchenko National University of Kyiv, Ukraine

Theory of queues is a key tool in modelling and performance analysis. But models of general form are extremely difficult to study. It leads to the need to simplify the models, to develop approximate methods for their study.

In this paper we consider a queueing network consisting of r service nodes. From outside a nonstationary Poisson flow of calls $\nu_i(t)$ with the leading function $\Lambda_i(t)$, $i=\overline{1,r}$ arrives to the i-th node. Each of r nodes operates as a multi-channel queueing system. If a call arrives at such a system then it is processed immediately. We study networks with different distribution type of service times at the nodes. Once the service is completed at the i-th node, the call is transferred to the j-th node

with probability p_{ij}, or it leaves the network with probability $p_{ir+1} = 1 - \sum_{j=1}^{r} p_{ij}$. Denote by $P = \| p_{ij} \|_1^r$ the switching matrix of the network. Define the service process in the network as an r-dimensional process $Q(t) = (Q_1(t),...,Q_r(t))'$, $t \geq 0$, where $Q_i(t)$, $i = \overline{1,r}$ is the number of calls at the i-th node at instant $t \geq 0$. The process $Q(t)$, $t \geq 0$, is analyzed provided that the network operates in heavy traffic regime. It is proved that under heavy traffic conditions the process $Q(t)$ can be approximated by a Gaussian process. Characteristics of limit processes are found. Justification is given as a functional limit theorem.

Keywords: Multi-channel Queueing Network, Nonstationary Poisson Input Flow, Gaussian Approximation.

Join Moments for the Backward and Forward Recurrence Times

Losidis Sotirios, Konstadinos Politis
Dept of Statistics and Insurance Science, Univ. of Pireus,Greece

We will study the join moment of the backward and forward recurrence time under the assumption that the distribution of the inter arrival times belongs to IMRL class. We will also present an upper bound for the k-th moment of the forward recurrence time and we will discuss the monotonicity behaviour for the second moment and for the variance of the forward recurrence time. Finally we will study the asymptotic covariance between the forward recurrence time and the number of renewals in a ordinary renewal process.

Keywords: renewal function, renewal density, mean forward recurrence time, IMRL, HNBUE, HNWUE.

Scan Statistics for Disease Clusters with Risk Adjustments

Wendy Lou
Dalla Lana School of Public Health, University of Toronto, Canada

The use of scan statistics for identifying clusters in biomedical research and public health sciences has become increasingly common.

Motivated by a cohort study to investigate the relationship between the health status of study subjects and their exposure to air pollutants (e.g. traffic pollution), we consider various approaches for analysis and focus on scan statistics in this presentation. In an attempt to adjust for risk factors, such as smoking status and obesity, a scan-like statistics will be developed and applied to the real application. Numerical comparisons and their interpretations will be presented to illustrate the results.

Keywords: Spatial Statistics, Pattern Classification, Heterogeneity.

Using an Extended Tobit Kalman filter in Order to Improve the Motion Recorded by Microsoft Kinect

K. Loumponias[1,2], N. Vretos[2], P. Daras[2], G. Tsaklidis[1]
[1,2]Dept of Mathematics, Aristotle University of Thessaloniki, Greece,
[2]ITI Centre for Research and Technology, Greece

In order to analyze data and improve the motion recorded by Microsoft Kinect 2 camera we apply an extended Tobit Kalman Filter methodology by assuming appropriate constrains in the related state equations. The data concern three-dimensional spatial coordinates recording movements of a persons' joints, which are subject to measurement errors. Simulations of skeleton motion before and after using the aforementioned extended Tobit Kalman Filter are presented.

Sir Ronald Fisher and Andrey N. Kolmogorov: An Uneasy Relationship

Mikhail Malyutov
Dept of Mathematics, Northeastern University, USA

During my work as academic secretary of the Kolmogorov Statistical Lab in Moscow University and in subsequent years I became interested in the history of Statistics in the former Soviet Union and the leading role of Kolmogorov in it as complemented to his world leadership in the Probability development. I'll discuss the facts not covered in my remarks published in Kolmogorov's selected works.
Unlike Probability, he had a great rival whose contributions in Statistics and Population Genetics were regarded as a beacon in Kolmogorov's research, notably the early notion of information.
Apparently the first fundamental Kolmogorov's contribution to the Mathematical statistics was his work on the Kolmogorov-Smirnov and related nonparametric tests which initiated the functional approach in statistics. Unfortunately, the Fisher's attitude was negative due to his

previously developed Permutation tests which he regarded as solely appropriate. This damaged feelings of his young rival. Partly due to this, Kolmogorov published a huge critical survey of ANOVA methods published in the Proceedings of the all-union conference on Statistics in Tashkent, 1949 and reproduced with minor changes in the fundamental textbook of Mathematical Statistics by V. Romanovsky.
It must be emphasized that 1949 was critical for the survival of Soviet Mathematical Statistics which has been under governmental attacks while applied Statistics (including the genetic applications developed by Kolmogorov) has already was prohibited in the USSR. The leading Soviet Probabilists published political and philosophical insinuations addressed to the leading western researchers (except for Kolmogorov and few others) in the Proceedings mentioned above.
My personal reflections about Kolmogorov's admiration of the Fisher's results will be included in my talk.

Sparsity Against Exponential Complexity in Big Data: Quality Improvement through Design & Data-Driven Modelling and Inference for Time Series

Mikhail Malyutov
Dept of Mathematics, Northeastern University, USA

1. SCOT models: studying sparsity of memory for time series modelling. Until recently, the statistical dependence between neighbours in discrete time series (`memory') was mainly modelled by ARIMA-type models (known as the Box-Jenkins technique in quality control), and related GARCH models for volatility of financial time series. Both approaches are isotropic–they use the same coefficients and order of the model irrespective of actual values of preceding measurements. This isotropy is inadequate in many applications, such as industrial, linguistic, financial, etc. Stochastic COntext Tree (abbreviated as SCOT) is m-Markov Chain with every state of a string independent of the symbols in its more remote past than the context of length determined by the preceding symbols of this state.
In all of explored applications we uncover a complex sparse structure of memory in SCOT models that allows excellent discrimination power. In addition, a straightforward estimation of the stationary distribution of SCOT gives insight into contexts crucial for discrimination between, say, different regimes of financial data or between styles of different authors of literary texts. We prove the Locally Asymptotic Normality of this model, Local Asymptotic Minimaxity of the Likelihood-based estimators and Locall Uniformly Maximal Power of the Likelihood-based tests.

2. We will also discuss the change of paradigm in the celebrated R. Fisher's combined ideas of randomization and Complete Factorial Designs CFD(d) of dimension d and the so-called Response Surface methodology under sparsity assumption. The random sample of the CFD(d) points and separate testing of inputs is shown to be asymptotically optimal to screen out active inputs in factorial models. Thus, sparsity is shown to be a key idea for fighting exponential complexity of big data as is also demonstrated in many works of D. Donoho and his collaborators.
3. All theoretical results are supported by intensive statistical simulation and analysis of real data.

Electricity Spot Price Modelling using a Higher-Order HMM

Rogemar Mamon, Heng Xiong
Dept of Statistical and Actuarial Sciences, University of Western Canada

A model for electricity spot price dynamics exhibiting stochasticity, mean reversion, spikes and memory is proposed. The parameters are modulated by a higher-order hidden Markov chain in discrete time, which captures the switching between different economic regimes resulting from the interaction of various factors. Adaptive filters are established and they provide optimal estimates for the state of the Markov chain and related quantities of the observation process. Estimated values of the model parameters are given in terms of the recursive filters. Our self-calibrating model is tested on a deseasonalised series of daily spot electricity prices compiled by Alberta Electric System Operator. Some implications of the model to the pricing of electricity contracts are explored.

Keywords: change of measure, EM algorithm, commodity derivatives, valuation, statistical estimation.

Random Utility Models in a Stated Preference Approach. Some Considerations on the Transition from University to Work

Paolo Mariani[1], Andrea Marletta[1], Mariangela Zenga[2]
[1]Dept of Economics Management and Statistics, University of Milano-Bicocca, Italy, [2]Dept of Statistics and Quantitative methods, University of Milano-Bicocca, Italy

Individual choice behavior involves relevant decisions among alternatives, measured on all type of scales and pertaining various research field, as labor force participation, family size, type of education etc. These observations can be obtained in preference elicitation experiments, in the context of stated. i:e: openly expressed, choice. Through conjoint measurement on two or more characteristics, their levels are ordered in a relation of preference between choice alternatives. Random utility models (RUMs) apply to the stated preference approach, maximizing a utility function that incorporates the simultaneous effects of two or more factors.

Let S and T be two non empty sets. A relation R from S to T is defined as a subset of the product set $S \times T$. A $S \times T$ relation is then the set attributes for an ordered pair of components, conjointly measured. It is possible to combine relations with a composition rule, $R_1 \circ R_2$. Let the sets S and T be the two nonempty sets:

$$S = (A, B, \dots, C, D); T = (Q, R, \dots, V, Z) \quad (1)$$

A utility function U is a function that maps from a finite set of alternatives X into the real field R, so that $x \preccurlyeq y \iff u(x) \preccurlyeq u(y)$. The utility function is a measure on an interval scale moving from the ordered preferences. The utility of the $j-th$ alternative $(j = 1, \dots, J)$ for the $i-th$ subject $(i = 1, \dots, n)$ can be partitioned into a systematic component V_{ij} and a random component ε_{ij}. ∨ :

$$U_{ij} = V_{ij} + \varepsilon_{ij};\ \forall i, i = 1, 2, \dots, n;\ \forall j, j = 1, 2, \dots, J \quad (2)$$

Multinomial Logit Model (MLM) is applied to estimate probabilities and related utilities, in discrete choice models, on the assumption that the error terms are multivariate normal or independently and identically Type I extreme value (Gumbel) distributed (Fig. 1).

The application is related to a preference elicitation experiment at a university in Northern Italy. The paper poses some applicative considerations on the stated preference experiment, in a RUMs methodological framework.

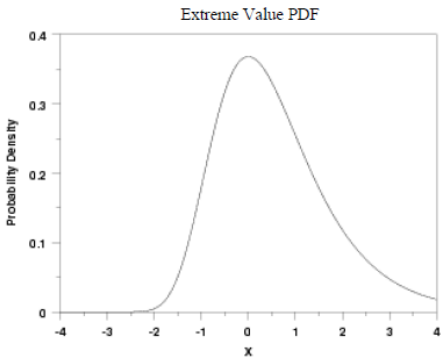


Fig. 1. Extreme value distribution

Keywords: Random Utility Models, Discrete Choice Models, Students, University.

Extremes in Random Graphs Models of Complex Networks

Natalia Markovich
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Russia

Regarding the analysis of social, Web communication and complex networks the fast finding of most influential nodes in a network graph constitutes an important problem. We consider two indices of the influence of those nodes, namely, PageRank and a Max-linear model. The latter is obtained by a substitution of sums in Google's definition of PageRank by maxima. Regarding the PageRank random walk renewals occur with probability 1-c (c is a damping factor) with time units delay after a visit with probability c to descendants of the underlying root node. means also the in-degree to node i. By its definition the PageRank of a Web page is the probability to find a random walk at this node when the process has reached the steady state. From another perspective, the PageRank of a randomly selected node is a Markov process due to the random number of in- and out-degrees of the node. We obtain the extremal index of both a PageRank and a Max-linear model. To this end, the PageRank process is considered as an autoregressive process with random coefficients. Those depend on ranks of incoming nodes and their out-degrees. Assuming that the coefficients and a noise are independent and distributed with regularly varying tail with the same tail index but with different values of extremal index it is proved that the extremal index is the same for both PageRank and the Max-linear model. In order to find the extremal index, we use the representation of the PageRank as Galton-Watson branching process with a specification. The nodes of the branching tree may have outbound stubs (teleportations) to arbitrary nodes of the network. Such graph is called a Thorny Branching Tree. We find the extremal index of the Max-linear model depending on the reproduction law assuming the independence of noise and using the fact that ranks of dangling nodes are determined only by noise due to the lack of incoming nodes.

Keywords: Extremal index, PageRank, Max-Linear model, Branching process, Autoregressive process, Tail index, Complex Networks.

Modelling Patient Waiting Time in Emergency Departments using Coxian Phase-type Distributions

Adele H. Marshall[1], Laura M. Boyle[1], Mark Mackay[2]
[1]Centre for Statistical Science and Operational Research (CenSSOR), Queen's University Belfast, Northern Ireland, U.K., [2]Health Care Management, School of Medicine, Flinders University, South Australia

There is a constant strain being placed on hospital emergency departments (EDs) around the world to provide urgent hospital care to increasing numbers of patients. In Australia, all public hospital EDs are subject to the National Emergency Access Target of treating or admitting patients within four hours. In the UK the four-hour target has recently been reviewed to only apply to urgent health problems as resources are not in place to apply to all patients arriving to emergency departments. Previous research has demonstrated how the Coxian phase-type distribution, a special type of Markov model which represents the time until absorption of a finite Markov chain in continuous time as a series of latent phases, can be used to represent time in hospital as a series of stages. The aim of this research is to use Coxian phase-type distributions to model the patient waiting times in a South Australian emergency department. The characteristics of patients that progress to the later phases can be identified, revealing groups of individuals who consume a greater proportion of resources.

Keywords: Coxian phase-type distributions, Markov models, Emergency care.

An ANOVA-Type Procedure for Replicated Spatial and Spatio-Temporal Point Patterns

Jorge Mateu[1], Jonatan Gonzalez-Monsalve[1], Ute Hahn[2], Bernardo Lagos[3]
[1]University Jaume I of Castellon, Spain, [2]Aarhus University, Denmark, [3]University of Concepcion, Chile

Several methods to analyse structural differences between groups of replicated spatio-temporal point patterns are presented. We calculate a number of functional descriptors of each spatio-temporal pattern to investigate departures from completely random patterns, both among subjects and groups. We develop strategies for analysing the effects of several factors marginally within each factor level, and the effects due to interaction between factors. The statistical distributions of our functional descriptors and of our proposed tests are unknown, and thus we use bootstrap and permutation procedures to estimate the null distribution of

our statistical test. A simulation study provides evidence of the validity and power of our procedures. Several applications in environmental and engineering problems will be presented.

Keywords: K-function; Non-parametric test; Permutation test; Spatio-temporal point patterns; Subsampling.

Joint Models for Time-to-Event and Bivariate Longitudinal Data: a Likelihood Approach

Marcella Mazzoleni, Mariangela Zenga
Dept of Statistics and Quantitative Methods,
Università degli Studi di Milano – Bicocca, Italy

The joint models analyse the effect of a longitudinal covariate onto the risk of an event. They are composed of two sub-models, the longitudinal and the survival sub-model.
In this paper the focus is on the case in which the longitudinal sub-model is bivariate, considering more than one longitudinal covariate. For the longitudinal sub-model a multivariate mixed model can be proposed. Whereas for the survival sub-model, a Cox proportional hazards model is proposed, considering jointly the influence of both the longitudinal covariates onto the risk of the event.
The purpose of the paper is to implement an estimation method that is able to deal with the computational problem given by the introduction of other covariates and the increase of the number of parameters that must be estimated in a model that is already highly computationally demanding.
The proposed method of estimation is based on a joint likelihood formulation and it is the generalisation of estimation methods implemented for the univariate joint models.
The estimation of the parameters is based on the maximisation of the likelihood function achieved through the implementation of an Expectation-Maximisation (EM) algorithm.
In the M-step a one-step Newton-Raphson update is used, as for the estimation of some parameters is not possible to obtain closed-form. In addition also a Gauss-Hermite approximation is applied for some of the integrals involved.

Keywords: Joint Models, Multivariate Mixed Models, Joint Likelihood Approach.

Improving Rexpokit's Krylov Subspace Matrix Exponential Methods for Markov Processes

Meabh G. McCurdy, Karen J. Cairns, Adele H. Marshall
Centre for Statistical Science and Operational Research (CenSSOR), School of Mathematics and Physics, Queen's University Belfast, Northern Ireland, United Kingdom

The calculation of the exponential of a matrix is arguably the most widely used and most widely studied matrix function. It is required for applications such as modelling complex biological systems through Markov models or for modelling population growth. The computational run time of this matrix function is a major drawback which gets progressively worse as the dimensions of the matrix increases. The current methods in place to calculate the matrix exponential such as the scaling and squaring algorithm combined with the Padé approximation work well with small matrices but become problematic when dealing with large sparse matrices. Krylov subspace methods are an alternative method for calculating the matrix exponential. One of the advantages with Krylov methods are that they project a matrix of dimension *n* onto a small Krylov subspace of dimension *m*, where *m* is much smaller than *n*. This research aims to look at Krylov subspace methods for calculating the matrix exponential with focus on improving software packages that currently exist. Expokit is a software package that incorporates Krylov subspace methods for the matrix exponential, which has routines to deal with both small dense matrices and large sparse matrices. This software can be implemented in R through the use of Rexpokit, which is a wrapper function for the Expokit. This research will demonstrate an investigation aiming to achieve the correct balance between the time step size and the order of Krylov subspace needed to achieve optimal efficiency. Currently (Rexpokit 0.24.1), the implementation of the Krylov subspace methods in R doesn't show much promise, especially when dealing with large sparse matrices; it is two times slower than EXPM. This research has identified areas in the code where modifications can be made, such as the tolerance, to achieve both higher accuracy and faster run times which results in the code now being 30 times faster than EXPM. This implementation now has more user options to deal with different size matrices. The benefits of the newly purposed code will be demonstrated through one example application.

Keywords: Matrix Exponential, Krylov Subspace, Efficiency, Rexpokit.

Parameters Estimation of Stochastic Differential Equations using First Passage Times and Inverse Gaussian Law

Samia Meddahi, Khaled Khaldi
Dept of Mathematics, Faculty of Sciences, Boumerdes University, Algeria

This article consists in a new method - Generalized Passage Times - for estimating the parameters of stochastic differential equations. We estimate the two parameters of the Black-Scholes model in a financial times series using the first passage times and the inverse gaussian law. We compare the empirical results of the estimation and forecast obtained by the first passage time method and those obtained by the Generalized Passage Times.

Keywords: Estimation, Stochastic differential equations, Black-Scholes model.

Data Analysis of Nanoliquid Thin Film Flow over an Unsteady Stretching Sheet in Presence of External Magnetic Field

Prashant G Metri, Sergei Silvestrov
Division of Applied Mathematics, UKK, Mälardalen University, Sweden

A mathematical model is developed to analyze data for nanoliquid film flow over an unsteady stretching sheet in presence of external magnetic field. The flow problem within a nanoliquid film of an unsteady stretching sheet where the governing partial differential equations with the auxiliary conditions are reduced to ordinary differential equations with the appropriate corresponding condition via similarity transformations. The analytical solutions of the resulting ODEs are obtained, and the analytical solutions of the original problem are presented. The resulting non-linear ODEs are solved numerically using Runge-Kutta-Fehlberg and Newton-Raphson schemes. A relationship between film thickness β and the unsteadiness parameter S is found. Besides, the effect of unsteadiness parameter S, the solid volume fraction of the nanoliquid ϕ, the Prandtl number Pr and the magnetic field parameter M, on the velocity and temperature distributions are presented and discussed in detail. Present data analysis shows that the combined effect of magnetic field and viscous dissipation has a significant influence in controlling the dynamics of the considered problem.

Keywords: Boundary layer flow, magnetic field, nanoliquid, thin film, similarity solutions, unsteady stretching sheet.

Bivariate Non-Central Polya-Aeppli Process and Applications

Leda D. Minkova
Sofia University, Bulgaria

In this paper we consider a stochastic process which is a sum of Poisson process and P'{o}lya-Aeppli process. The resulting process is called a noncentral P'{o}lya-Aeppli process (NPAP). The probability mass function, recursion formulas and some properties are derived. Then, by trivariate reduction method we introduce a bivariate non-central Polya-Aeppli process (BNPAP). As application we consider a bivariate risk model with BNPAP counting process. The ruin probability for the defined model is analyzed. As example we consider the case of exponentially distributed claims.

Keywords: Polya-Aeppli process, bivariate risk model, pure birth process, ruin probability.

The Flexible Beta Regression Model

Sonia Migliorati, Agnese M. Di Brisco, Andrea Ongaro
Dept of Economics, Management and Statistics
University of Milano-Bicocca, Italy

A relevant problem in applied statistics concerns modelling rates, proportions or, more generally, continuous variables restricted to the interval (0,1). The aim of this contribution is to study the performances of a new regression model for continuous variables with bounded support that extends the well-known Beta regression model (Ferrari and Cribari-Neto, 2004, Journal of Applied Statistics). Under our new regression model (Migliorati, Di Brisco and Ongaro, submitted paper), the response variable is assumed to have a Flexible Beta (FB) distribution, a special mixture of two Beta distributions that can be interpreted as the univariate version of the Flexible Dirichlet distribution (Ongaro and Migliorati, 2013, Journal of Multivariate Analysis). In many respects, the FB can be considered as the counterpart on (0,1) to the well-established mixture of normal distributions sharing a common variance. The FB guarantees a greater flexibility than the Beta distribution for modelling bounded responses, especially in terms of bimodality, asymmetry and heavy tails. The peculiar mixture structure of the FB makes it identifiable in a strong

sense and guarantees a likelihood a.s. bounded from above and a finite global maximum on the assumed parameter space.
In the light of these many theoretical properties, the new model results to be very tractable from a computational perspective, in particular with respect to posterior computation. Therefore, we provide a Bayesian approach to inference and, in order to estimate its parameters, we propose a new mean-precision parametrization of the FB that guarantees a variation independent parametric space. Interestingly, the FB regression model can be understood itself as a mixture of regression models.
Here we aim at showing the feasibility and strength of our new FB regression model by means of some simulation studies and applications on real datasets, with special attention to bimodal response variables and response variables characterized by the presence of outliers. To simulate values from the posterior distribution we shall implement the Gibbs sampling algorithm through the BUGS software.

Keywords: Beta regression, Flexible Dirichlet, Mixture models, Proportions, MCMC.

Stochastic Models for Biological Populations with Sexual Reproduction

Manuel Molina, Manuel Mota, Alfonso Ramos
Dept of Mathematics, University of Extremadura, Spain

Branching model theory provides appropriate mathematical models to describe the probabilistic evolution of dynamical systems whose components after certain life period reproduce and die in such a way that transition from one to other state of the system is made according to a certain probability law. Nowadays, this theory is an active research area of both theoretical interest and applicability to such fields as biology, demography, ecology, epidemiology, genetics, population dynamics, and others. Most biological species reproduce sexually which requires the involvement of females and males in the population. Moreover two important phases are carried out: mating and reproduction. We focus the interest on the development of stochastic models to describe the demographic dynamics of biological populations with sexual reproduction. In the last years, this research line has received several theoretical and applied contributions. In particular, new classes of two-sex branching models where, in each generation, the reproduction phase is developed in a random environment, or where both biological phases, mating and reproduction, are influenced by the current numbers of females and males in the population, have been investigated. In this talk, we will present several methodological results concerning such classes

of two-sex models. As illustration, we will show some simulated examples.

Keywords: Stochastic modeling, Branching models, Two-sex models, Population dynamics.

Acknowledgements: *This research is supported by the Junta de Extremadura (GR15105), the Ministerio de Economía y Competitividad (MTM2015-70522-P), the National Fund for Scientific Research (Bulgaria, DFNI-I02/17) and the FEDER.*

Bayesian Multidimensional Item Response Theory Modeling Using Working Variables

Alvaro Montenegro, Luisa Parra
Departamento de Estadistica, Universidad Nacional de Colombia, Colombia

We propose a hybrid Metropolis-Hastings within Gibbs type algorithm with independent proposal distributions to the latent traits and the item parameters in order to fit classical Multidimensional Item Response Theory Models. The independent proposals are all multivariate normal distributions based on the working variables approach applied to the latent traits and the item parameters. The covariance matrix of the latent traits is estimated using an inverse Wishart distribution.
The results show that the algorithm is very efficient, effective, and yields to high acceptation rates. The algorithm is applied to real data from a large-scale test applied in Universidad Nacional de Colombia.

Keywords: Multidimensional Item Response Theory, working variables, Bayesian Modeling, Large scale tests.

Lie Symmetries of the Black – Scholes Type Equations in Financial Mathematics

Asaph K. Muhumuza, Silvestrov Sergei, Anatoliy Malyarenko
Division of Mathematics, The School of Culture and Communication (UKK), Mälardalen University, Sweden

Stochastic differential equations are strongly linked to partial differential equations. Lie groups and Lie algebras are a strong tool in analysis of global solutions of partial differential equations that occur in most of mathematical models in finance. We give the symmetry analysis of a couple of popular financial models.

Keywords: Lie groups, Lie algebras, Lie symmetries of differential equations, Black-Scholes type model.

Greece and India the Countries of Great Heritages: Facing Critical Socio-Economic Crisis

Barun Kumar Mukhopadhyay
Population Studies Unit, Indian Statistical Institute, India

Greece a country of great social, cultural and political heritage with oldest democracy even started prior to the 2nd millennium BC now has widespread and perpetual news circulated through different print and electronic media over her worst economic hardship that the people are paying several taxes with acute unemployment problem among young graduates, in particular. As par Lois Lambrianidis (2016), a professor at Macedonia University, Thessaloniki, for example, 190,000 Greek scientists are currently working abroad today, mainly because of the limited demand from the Greek economy for university graduates. India also a country with similar great social, cultural and political heritage with largest democracy has also been passing through similar crisis. These scenarios of Greece and India has forced social scientists, other research scholars, planners, executors and many other NGOs to take up research study. The present paper attempts to analyse the socio-demographic-economic-infrastructural aspects of the population of Greece and India in order to have a perspective scenario. When Greece's economic crisis might have started a couple of years back, India's situation started too early period almost after independence in 1947. The findings, to some extent show a moderate standard of life including high literacy rate with negative growth of population during 2001-2011 census in Greece, but in India, there is low literacy rate particularly of females with slightly declining population growth of the order of 2.16 per cent in 2001 to 1.97 per cent in 2011 census. There are many other aspects remain to be investigated. Though the two countries have large differences in population when India's population stood at 1210 million in 2011, Greece's population stood only at 11 million (approx) in the same census year, still their problems are more or less similar. It is also to investigate the effects of two recent international events firstly BREXIT and the ensuing probable American new policies over the two countries.

Keywords: Greece, India, NGOs, democracy.

"Big Data" Triangle and Modern Data Analysis

Subhadeep ("Deep") Mukhopadhyay
Department of Statistical Science, Temple University, U.S.A.

"Difficulties in identifying problems have delayed statistics far more than diffculties in solving problems. This seems likely to be the case in the future, too." John Tukey.

An important class of learning problem that has recently attracted researchers from various disciplines including neuroscience, theoretical computer science, information theory, statistical physics, genomics, and machine learning shares a common interesting structure: massive number of data points are collected from a vast collection of distributed data sources and sensors at a very fast speed from a probability distribution $\boldsymbol{p} = (p_1, \dots, p_k)$ over a large domain k, where $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$.

We call this new frontier as `Digital Big Data Processing,' by drawing a historical parallel with the 1960s Digital (or Discrete) Signal Processing movement. The key focus of our research program is to design an efficient learning model to comprehensively analyze this typical data structure, allowing compression and fast computation (time and storage-efficient).

We will present a new viewpoint that might give us taxonomy way of thinking about this general research field on fast approximate computing and statistical learning (where Learning = Modeling + Inference) of digital big data. Central to our approach is a new functional representation scheme (nonparametric harmonic analysis result) to reduce the size of the problem (dimension of the sufficient statistics) with an eye towards understanding and developing efficient algorithms that are fast enough and accurate enough to make a big difference. Our modeling approach works convincingly well in practice and often outperforms the recent `breakthrough' algorithms (of theoretical computer science) by a handsome margin.

Keywords: Nonparametric modeling, large sparse distribution modeling, Data-effitient learning, and LP Orthogonal System.

References:

[1]. Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author) Statistical Science 16 (3), 199-231.

[2]. Canonne, C. L. (2015). A survey on distribution testing: Your data is big. but is it blue? Electronic Colloquium on Computational Complexity (ECCC) 22 (63).

[3]. Khmaladze, E. (2013). Note on distribution free testing for discrete distributions. The Annals of Statistics, 41 2979-2993.

[4]. Mukhopadhyay, S. (2016). Large scale signal detection: A unifying view. Biometrics, 72 325-334.

[5]. Mukhopadhyay, S. (2017). Large-scale mode identification and data-driven sciences. Electronic Journal of Statistics, 11 215-240

[6]. Mukhopadhyay, S. and E. Parzen (2014). LP approach to statistical modeling. Technical Report available at arXiv:1405.2601 .
[7]. Orlitsky, A., N. P. Santhanam, and J. Zhang (2003). Always good turing: Asymptotically optimal probability estimation. Science 302 (5644), 427-431.
[8]. Parzen, E. and S. Mukhopadhyay (2013a). United Statistical Algorithms, LP comoment, Copula Density, Nonparametric Modeling. 59th ISI World Statistics Congress (WSC), Hong Kong.
[9]. Parzen, E. and S. Mukhopadhyay (2013b). United Statistical Algorithms, Small and Big Data, Future of Statisticians. Technical Report available at arXiv:1308.0641 .
[10]. Rubinfeld, R. (2012). Taming big probability distributions. XRDS: Crossroads, The ACM Magazine for Students 19 (1), 24-28.
[11]. Simonoff, J. S. (1985). An improved goodness-of-fit statistic for sparse multinomials. Journal of the American Statistical Association, 80 671-677.
[12]. Zelterman, D. (1987). Goodness-of-fit tests for large sparse multinomial distributions. Journal of the American Statistical Association, 82 624-629.

Population Explosion: Challenges in Management in the Megacities of India

K. Shadananan Nair
Nansen Environmental Research Centre (India), India

Uncontrolled migration leads to spreading of slums, resulting in environmental degradation in the megacities – Mumbai, Delhi, Kolkata and Chennai - of India. Administration fails to provide basic facilities to the fast rising population. Things are worse in Mumbai where half of the population lives in slums. In terms of the population exposure to coastal environment, Mumbai and Kolkata are at high risk. Rural unemployment, industrial development, changing climate and the growing economic imbalance are prominent factors behind migration. Unwise planning and unscientific construction of sewerages create flooding, polluting entire water resources. Poor sanitation together with insufficient basic infrastructure creates serious health issues. Extremes in climate add to this. Escalating number of vehicles, inefficient traffic system, poorly maintained roads and encroachment of footpaths by street vendors create long hours of traffic jam. Industries and urban settlements do not have proper treatment. Rising population leads to several social issues such as conflicts over the allocation of water, food, energy and land. Population control and urban planning and management have become complicated, as they involve several socio-economic, environmental and political issues. Cities also face the issue of illegal migrants from neighbouring countries. There is a large number of unaccounted population and the national agencies fail to provide a reliable statistics. Present study analyses the factors behind the population explosion and the impact of increasing population on the megacities of India under a changing climate and environment and critically reviews the current policies and strategies. There are options to overcome the crisis such as

satellite cities with all basic facilities, increased rural employment opportunities, urban poverty eradication schemes and modernisation of the urban infrastructure to cope with the changing demographic and climate patterns. India needs an appropriate urban policy and population policy. Guidelines for this have been provided.

Keywords: megacity, population, India, migration, environment, climate.

Forecasting with Functional Data: Case Study

Laurynas Naruševičius, Alfredas Račkauskas
Dept of Mathematics and Informatics, Vilnius University, Lithuania

The question what is better, to forecast the aggregate quantity directly and then disaggregate, or to forecast the individual components and then aggregate them to form the forecast of the total is important in many applications. This is also known as top-down versus bottom-up forecasting problem. However, in any specific application usually it is difficult to argue on a theoretical ground what approach should be taken. This question usually is settled empirically. In our case study we consider the forecasting problem with functional data that represents the capital adequacy ratio and determines the capacity of the bank to meet potential losses arising from credit risk, operational risk and others. The aggregate forecast is important to macroprudential supervision whereas forecast of the individual institution is important to microprudential supervision. Since European banking sector is heterogeneous, first we perform a cluster analysis. The forecasting problem is solved by fitting functional regression models at individual component of a cluster and at aggregated level as well.

Forecasting with functional sample is used in many fields, e.g., energy market, finance, environment, mortality and fertility and others. Most of the studies are dealing with a temporal dependencies among functional data, i.e., functional autoregressive models are applied since usually the purposes are to predict possible paths of a random function. In our case the aim is to forecast the future of the particular random processes.

Keywords: Functional data, regression, forecasting.

Modelling non-Anticipated Longevity Shocks under Lee-Carter Model

Eliseo Navarro, Pilar Requena
Dept of Economics and Management Sciences, University of Alcalà, Spain

Under dynamic mortality models such as Lee-Carter (1992) we can observe that enlarging the sample period produces very significant changes in mortality rates forecasts due to unexpected shocks in deaths that affect both the shape and the level of the mortality surface predicted by these models. In this paper, we examine the behavior of mortality rates in Spain during the period 1975-2012 trying to analyze the effects of non-expected mortality rate changes and its impact on mortality rates forecasts derived of the implementation of Lee-Carter (1992) model. We propose a single factor model to capture these non-anticipated movements of the mortality surface consisting of identifying the key age that best explains the whole movement of mortality rates. Then we proceed to extrapolate these changes in the key mortality rate throughout the entire mortality surface. This model could be easily extended to a multifactor model. The resulting model can be used as powerful instrument for mortality risk management.

Keywords: Lee-Carter model, mortality shocks forecast, mortality risk management.

Context-Specific Independence in Technology Study

Federica Nicolussi, Manuela Cazzaro
Dept of Statistics and Quantitative Methods, University of Milano- Bicocca, Italy

In the field of the ordinal variables, with the term context-specific independence we refer to the particular conditional independence that holds only for some modalities of the variable in the conditioning set but not for all. That is, given three variables A, B and C we describe this situation as $A \perp\!\!\!\perp B | C=c$, where c is a subset of all possible values of C. Nyman (2016) dealt with the non-ordinal variables. In this work, we present a model able to capture this situation by using the marginal log-linear parameterization and ordinal variables. The need of this parameterization chases the will of consider simultaneously marginal and conditional independencies. This parameterization uses local, global or continuation logits evaluated on different marginal contingency tables in order to consider the ordered modalities of the variables. For this reason, these logits are appropriate also to explain the contribution of a modality of the conditioning variables on the other ones. This model is

also represented through a Stratified Chain Graphical Model (SCGM), an extension of SGM proposed by Nyman (2016), that use a chain graph model to represent the classical conditional independencies and labeled arcs in the graph to denote context-specific independencies.
The proposed model is here used to analyze the "*Survey on information and communication technology (ICT) usage in enterprises*" of Istat (2014), in order to investigate the use of technologies in Italian enterprises. At this aim, we take into account different variables which can help us to define and study the phenomenon. The results of the analysis are represented by both a SCGM and a set of marginal log-linear parameters which together describe the relationship among the selected variables.

Keywords: Context-specific independencies, ordinal variables, graphical models, technology.

Probabilistic Modeling of Hydraulic Conditions in Pipeline Systems under a Random Set of Boundary Parameters at Nodes

Nikolay N. Novitsky[1,2], Olga V. Vanteyeva[1,2]
[1]Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Russia, [2]Skolkovo Institute of Science and Technology, Skoltech Center for Energy Systems, Russia

In practice hydraulic conditions of pipeline systems are calculated for two purposes: 1) to estimate pipeline capacity at specified (maximum, as a rule) loads of consumers and 2) to estimate an extent to which the consumers are provided with a target product at specified characteristics of consuming systems. The first (the main) type of calculations is based on the models with lumped loads of consumers and is applied at the stages of design, expansion, reconstruction of pipeline systems as well as when planning their main operating parameters. The second (checking) type of calculations is based on the application of models with non-fixed loads and is applied at the stage of pipeline system operation while calculating and analyzing off-design conditions, for example, emergency ones.
In both cases researchers traditionally apply deterministic models of flow distribution. However, the actual operating conditions of pipeline systems are formed under the influence of a great amount of random impacts of the external environment (consumer loads, pressure at sources, etc.). This explains the relevance of solving the problems of probabilistic modeling of steady-state hydraulic conditions to obtain the results in the form allowing their probabilistic interpretation.

In this paper consideration is given to a problem of probabilistic modeling of pipeline system hydraulic conditions which involves the models with lumped loads, and suggests setting pressure at more than one node (for example at the points where working medium comes). Thus, the nodal boundary conditions are specified, when either flow rate or pressure is given at each node. Being an extension to the previously proposed methods for probabilistic analysis of hydraulic conditions [1-3], such a statement of the problem has not been considered before, but it has an independent value and is of great applied relevance for practical design and operation of pipeline systems.

The paper substantiates the final formulas for the calculation of probabilistic means, variances and covariances of desired state variables on the basis of information on boundary conditions specified in a probabilistic form. The numerical example demonstrates the operability and high computational efficiency of the proposed approach compared to the traditional methods of Monte-Carlo type as well as greater applied value of the probabilistic methods compared to the deterministic ones.

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]. Novitsky N.N., Vanteyeva O.V. Modeling of stochastic hydraulic conditions of pipeline systems // Chaotic Modeling and Simulation (CMSIM). – 2014. – No.1. – P. 95-108.

Probabilistic Modeling of Hydraulic Conditions of Pipeline Networks under Random Composition of Boundary Conditions at Nodes

Nikolay N. Novitsky, Olga V. Vanteyeva
Energy Systems Institute SB RAS, Russia

The paper is concerned with the problem of probabilistic analysis of hydraulic conditions in pipeline networks, which occur under the impacts of external environment. These impacts are taken into account to specify boundary conditions for nodal flow rates or pressures in a probabilistic form. The research reveals practical value of such a statement which arises at the stages of design and operation of pipeline networks in the analysis of their transmission capacities and feasibility of operating conditions. A mathematical statement and a general scheme for solving the problems are presented. The final relationships are obtained to calculate the mean value and covariance matrices of the sought state variables. The relationships provide analytical representation of the model of probabilistic flow distribution. A numerical example is presented

to illustrate high computational efficiency of the proposed method for probabilistic analysis of operating conditions and its advantages over the traditional deterministic models and methods for such an analysis.

Keywords: Pipeline systems, probabilistic modeling, flow distribution, hydraulic circuits, statistical parameters.

A comprehensive Study of Lattice Pricing beyond Black and Scholes

Carolyne Ogutu[2], Karl Lundengård[1], Ivivi Mwaniki[2], Sergei Silvestrov[1], Patrick Weke[2]
[1]Division of Applied Mathematics, UKK, Mälardalen University, Sweden,
[2]School of Mathematics, University of Nairobi, Kenya

Derivatives can be defined as an agreement based on an underlying asset or a non-tradeable asset. Options are derivatives that give the holder the right but not obligation to exercise it before or at maturity. The seminal model by Black and Scholes doesn't reflect real asset dynamics due to its continuous trading and constant volatility assumptions. Cox et al first developed lattices – binomial lattice - as a numerical scheme that discretizes the lifespan of the option by dividing it into time steps of equal length. Dynamic programming is then used to obtain the option price at inception. Other lattice schemes have been developed such as trinomial and pentanomial. In this paper, we undertake a comprehensive study of lattice models pre (Cox et al and family) and post (Amin and family) Black Scholes model. In addition, we develop an extended lattice-pricing model, which enhances the assumptions by post-Black Scholes lattice models.

Keywords: Lattice models, moment-matching, Vandermonde matrix, option pricing.

Extreme Points of Ordinary and Generalized Vandermonde Determinants

Jonas Österberg, Sergei Silvestrov, Karl Lundengård
Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Sweden

The Vandermonde determinant has some interesting geometric and algebraic properties as a multivariate function. Due to its symmetry, optimization over symmetric surfaces in various dimensions will lead to symmetrically placed solutions, and in many cases these solutions are most easily constructed as the roots of some family of polynomials. In

this paper we explore the ordinary and generalized Vandermonde determinant and these, often orthogonal, families of polynomials. We also consider the behavior of these determinants over other bounded and unbounded surfaces of interest.

Keywords: Vandermonde determinant, Optimization, Orthogonal polynomials.

Saturn Rings: Fractal Structure and Random Field Model

Martin Ostoja-Starzewski[1], Anatoliy Malyarenko[2]
[1]Dept of Mechanical Science & Engineering, Institute for Condensed Matter Theory and Beckman Institute, University of Illinois at Urbana-Champaign, U.S.A.,
[2]Division of Applied Mathematics, Mälardalen University, Sweden

This study is motivated by a recent observation [1], based on photographs from the Cassini mission, that Saturn's rings have a fractal structure in radial direction. Accordingly, two questions are considered: (1) What Newtonian mechanics argument in support of that fractal structure is possible? (2) What kinematics model of such fractal rings can be formulated? Both challenges are based on taking Saturn's rings' spatial structure as being statistically stationarity in time and statistically isotropic in space, but statistically non-stationary in space. An answer to the first challenge is given through the calculus in non-integer dimensional spaces and basic mechanics arguments [2]. The second issue is approached in Section 3 by taking the random field of angular velocity vector of a rotating particle of the ring as a random section of a special vector bundle. Using the theory of group representations, we prove that such a field is completely determined by a sequence of continuous positive-definite matrix-valued functions defined on the Cartesian square F^2 of the radial cross-section F of the rings, where F is a fat fractal.

Keywords: Saturn rings, fractal, dynamics.

References:

[1]. J. Li and M. Ostoja-Starzewski (2015), Edges of Saturn's rings are fractal, *SpringerPlus*, **4**, 158. *arXiv: 1207.0155* (2012).

[2]. V.E. Tarasov (2006), Gravitational field of fractal distribution of particles, *Celest. Mech. Dyn. Astron*. **94**, 1-15.

Tensor Random Fields in Conductivity and Classical or Microcontinuum Theories

Martin Ostoja-Starzewski[1], Anatoliy Malyarenko[2]
[1]Dept of Mechanical Science & Engineering, Institute for Condensed Matter Theory and Beckman Institute, University of Illinois at Urbana-Champaign, U.S.A., [2]Division of Applied Mathematics, Mälardalen University, Sweden

We study the basic properties of tensor random fields (TRFs) of a wide-sense homogeneous and isotropic kind with generally anisotropic realizations. Working within the constraints of small strains, attention is given to anti-plane elasticity, thermal conductivity, classical elasticity and micropolar elasticity, all in quasi-static setting albeit without making any specific statements about the Fourier and Hooke laws. The field equations (such as linear and angular momentum balances and strain-displacement relations) lead to consequences for the respective dependent fields involved. In effect, these consequences are restrictions on the admissible forms of the correlation functions describing the TRFs.

Keywords: elasticity, conductivity, microcontinuum, tensor random fields, correlation functions

References:

[1]. M. Ostoja-Starzewski, L. Shen and A. Malyarenko (2015), Tensor random fields in conductivity and classical or microcontinuum theories, *Math. Mech. Solids* **20**(4), 418-432.

A New Distribution for the Fatigue Lifetime

Gamze Ozel[1], Selen Cakmakyapan[2]
[1]Dept of Statistics, Hacettepe University, Turkey,
[2]Dept of Statistics, Istanbul Medeniyet University, Turkey

Fatigue is the weakening of a material is exposed to stress and tension vacillations. When a metarial subjected to cycling loading, fatigue occurs and it is a progressive and localised structual damage. The fatigue process (fatigue life) begins with an imperceptible fissure and is growth, propagated by the cyclic patterns of stress on the material. Consequently, this process causes the rupture or failure of this material. The failure occurs when the total extension of the crack exceeds a critical threshold for the first time. The partial extension of a crack produced by fatigue in each cycle is modeled by a random variable. The variable depends on a lot of factors such as the type of material, the number of previous cycles, the magnitude of the stress. Birnbaum and Saunders (1969) proposed a two-parameter distribution to model failure time due to fatigue. In this paper, we defined a new three-parameter fatigue lifetime model called Lindley Birnbaum and Saunders (LBS). We obtained characteristical functions and properties of the distribution.

Also, a real data set was fitted by LBS and some known distributions. We showed LBS distribution is the best fit among the others.

Keywords: Fatigue lifetime, Birnbaum Saunders Distribution, Lindley Distribution.

Odd Log-Logistic Power Lindley Distribution with Theory and Lifetime Data Application

Gamze Özel Kadılar[1], Emrah Altun[1], Morad Alizadeh[2]
[1]Dept of Statistics, Hacettepe University, Turkey,
[2]Dept of Statistics, Faculty of Sciences, Persian Gulf University, Iran

The statistical analysis and modeling of lifetime data are essential in almost all applied sciences including, biomedical science, engineering, finance, and insurance, amongst others. A number of one parameter continuous distributions has been introduced in statistical literature including exponential, Lindley, gamma, lognormal, and Weibull. The exponential, Lindley and the Weibull distributions are very popular. The Lindley distribution is a very well-known distribution that has been extensively used over the past decades for modeling data in reliability, biology, insurance, finance, and lifetime analysis. It was introduced by Lindley (1958) to analyze failure time data. However, the need for extended forms of the Lindley distribution arises in many applied areas.
One parameter Lindley distribution does not provide enough flexibility for analyzing different types of lifetime data. Hence, it will be useful to consider other alternatives to this distribution for modelling purposes.
The goal of the present study is to introduce a new distribution using the power Lindley distribution as the baseline distribution. In this study, we derive various of its structural properties including ordinary and incomplete moments, quantile and generating functions and order statistics. The new density function can be expressed as a linear mixture of exponentiated Lindley densities. The maximum likelihood method is used to estimate the model parameters. Simulation results to assess the performance of the maximum likelihood estimation are discussed. We prove empirically the importance and flexibility of the new model in modeling lifetime datasets.

Keywords: Lindley distribution, odd log-logistic generalized family, moments, maximum likelihood.

Prospective Scenarios of Death Coverage of the Northeast Brazil

Neir Antunes Paes[1], Alisson dos Santos Silva[2]
[1]Postgraduate Program in Decision Modelling and Health, Federal University of Paraíba, Brazil, [2]Postgraduate in Mathematical and Computational Modeling, Federal University of Paraíba, Brazil

Vital statistics reflect the health status of a population, which are widely used in the formulation of important demographic indicators. The evolution of vital records in Brazil is marked by political factors and administrative instabilities that have compromised its quality and utility. Due to this commitment, the two main sources of vital records, the Brazilian Institute of Geography and Statistics and the Ministry of Health do not capture all of these records, mainly in less developed regions such as the Northeast of Brazil with a population of 56 million inhabitants in 2016. Although there have been gradual advances in coverage of deaths in Brazil, the Northeast region has not yet reached universalization (100%). Among the nine states that compose this region, coverage of deaths in 2011 ranged from 79-94%. In order to estimate the year in which the states of the Northeast will reach the universalization of death records projections were performed on coverage of deaths for each state. The annual series of death coverage estimated by the Ministry of Health from 1991 to 2011 were used. The projections were made through the mathematical methods of projections: Logistic, Gompertz and Holt's Exponential Smoothing Model. The model of Holt, in general, was the best fit to the pattern of the series of coverage of deaths. The states were classified in three intervals of years when they reached 100% of coverage, which varied from 2019 to 2028. It is estimated that for the Northeast the universalization of deaths will be reached around 2021. It is expected that these scenarios can contribute to the planning strategies and to the evaluation of managers regarding the actions and policies to be implemented and executed on the performance of death statistics in the Northeast and Brazil.

Keywords: Vital Statistics, Mortality, Death Coverage, Brazil.

The Variances and Covariances of the Macro State Sizes via the Micro State sizes in semi Markov modeling for a Healthcare System

Aleka Papadopoulou[1], Sally McClean[2], Zacharias Kyritsis[1], Lalit Garg[3]
[1]Department of Mathematics, Aristotle University of Thessaloniki, Greece, [2]School of Computing and Information Engineering, University of Ulster, Northern Ireland, UK, [3]Computer Information Systems, Faculty of Information & Communication Technology, University of Malta, Malta

We have previously discussed how a semi Markov model can be converted to a Markov model with new (micro) states corresponding to a composite of the original state and duration. This approach can be applied to a healthcare system comprising of patients who move though different states of care. Thus we can derive the expected population structure and the variances and covariances of the macro state sizes of a semi Markov system via the micro state sizes of the Markov system. Healthcare data of stroke patients are applied to illustrate the previous theoretical results.

Keywords: Semi-Markov systems, Variances, Covariances, Healthcare.

Estimation of a Two Variable Second Degree Polynomial

Papatsouma Ioanna, Farmakis Nikolaos, Ketzaki Eleni
Dept of Mathematics, Aristotle University of Thessaloniki, Greece

In various fields of environmental and agriculture sciencesthe estimation of a two variable 2^{nd} degree polynomial coefficients via Sampling is of major importance, as it gives very useful information. In this paper, we propose a very simple and very low budget systematic sampling plan for the estimation of the coefficients $A,\ B,\ C,\ D,\ E$ and H of the polynomial $f(x,y) = (Ax^2 + By^2 + Cxy + Dx + Ey + H)^{-1}$, which is sometimes found to be a probability density function. The above polynomial is defined on a domain $D = [a,b]\mathrm{x}[\mathrm{c},\mathrm{d}]$, which can be represented by the domain $D = [0,1]x[0,1]$ for convenience. Numerical methods, such as Simpson's rule, are applied. The comparison between means of both estimated and theoreticfunctions is used to confirm the accuracy of the results. The stability of the numerical methods allows us to get results with very good accuracy for small sample sizes. Illustrative examples are given.

Keywords: Systematic sampling, polynomial, coefficients, Simpson.
MSC2010 Classification: 62D05, 62E17

Employers' Assessments on Hiring: Results from a Vignette Experiment

Dimitris Parsanoglou, Aggeliki Yfanti
Dept of Social Policy,
Panteion University of Social and Political Sciences, Greece

The purpose of this paper is to analyse employers' assessments on hiring new employees. Based on a factorial survey experiment, where respondents who make hiring decisions, i.e. employers or human resources managers, it focuses on analysing hiring decisions taking into account gender, age, educational level, qualification, unemployment spells during the early professional life. The analysis is based on the data from an employer vignette experiment conducted in Greece in five occupational fields: mechanics, health, restaurant-services, IT and finance. The survey focused on the transition from education to the labour market, i.e. designed vignettes represented hypothetical CVs of young candidates in their early career.

The main hypotheses interrogate whether unemployment spells during the early professional life have any signalling effect on young people. The case of Greece in particular, where youth unemployment rate is still almost 50%, can be revealing as whether bad macro-economic conjuncture can produce a decreasing stigmatisation of those who are left out of employment for a certain, even long, period of time. Another interesting hypothesis that can be tested through multi-level models and random effect analysis, is the comparability between different sectors; especially, when it comes to different requirements of skills and expertise, different educational levels and, even more importantly, different patterns of hiring procedures, e.g. in health sector large part of employers were households whereas in IT or finance were businesses.

The paper will provide an overview of the most significant factors that interplay in employers' selecting decisions, interrogating whether unemployment in comparison to other variables constitutes an important source of stigma or crisis and high unemployment mitigates stigmatisation, rendering other variables more important.

Keywords: Employers, hiring process, signalling effect, crisis.

Acknowledgements: *This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649395 (NEGOTIATE – Negotiating early job-insecurity and labour market exclusion in Europe, Horizon 2020, Societal Challenge 6, H2020-YOUNG-SOCIETY-2014, Research and Innovation Action (RIA), Duration: 01 March 2015 – 28 February 2018).*

Air Pollution Variability within an Alpine Italian Province

Giuliana Passamani[1], Matteo Tomaselli[2]
[1]*Dept of Economics and Management, University of Trento, Italy*
[2]*PhD in Economics and Management, School of Social Sciences, University of Trento, Italy*

As it's generally acknowledged that greenhouse gases and atmospheric aerosols represent a major causal force of climate change and that they may interact physically and chemically in the atmosphere and make it much more difficult to forecast the future variations in climate and in global warming, we can understand the importance of empirical modelling the time-series data measuring their levels. An important question is, therefore, understanding the temporal evolution of the gaseous and aerosol pollutants and how they interact among themselves and with atmospheric factors. Multivariate analysis of non-stationary but co-integrated time series data allows the identification of hidden deterministic and stochastic relations, thus contributing to an understanding of causal relationships in environmental problems. The data set is made up of daily time series observations on the main pollutants and meteorological variables, covering a period of fourteen years and recorded at different monitoring sites in an alpine Italian province. Data are characterized by trend-cycle and seasonal components that are to be taken into account in the analysis. The methodological approach we follow aims to detect within-province intra e inter-annual variability in air pollution concentrations in relation with meteorological conditions. In particular, the main purpose is the proposal of an analytical procedure that, moving from the statistical properties of the observed non-stationary time-series for each monitoring site, identifies the stochastic processes generating them and, addressing the complications inherent in statistical analyses of observational data, uncovers empirical relationships between and within the sites. The overall main aim is to assess whether any improvement in the pollution level has been detected during the period of observation. The results show that some improvement in the level of air pollution has been achieved even if there are evident differences among the monitoring sites.

Keywords: Air pollution, Time-series statistical properties, empirical modelling, spatial variability.

Redistricting Spain. A Proposal for an Unbiased System

Jose M. Pavía[1], Alberto Penadés[2]
[1]Elections and Public Opinion Research Group (GIPEyOP), Universitat de Valencia, Spain, [2]Elections and Public Opinion Research Group (GIPEyOP), Universidad de Salamanca, Spain

The Spanish election system is biased. Despite using proportional rules to both allocate seats among constituencies and, within each constituency, apportion seats among parties, the Spanish system does not treat equally all political forces. With the same number of votes nationwide, the system rewards conservative and oldest parties over progressive and newest parties: PP over PSOE; and, PP and PSOE over UP and C's.
Spain is divided in 52 constituencies that greatly vary in size and sociodemographic composition, which irrespectively of their populations distribute a minimum of two seats (except for the autonomous cities of Ceuta and Melilla where only one seat is allocated). The combined effect of these three issues provokes (in the same way than happens in another countries) the bias. Several proposals have been made trying to solve this bias, but they are made at the cost of losing the current virtues of the system, which are of great value. In this paper, we propose a solution respectful with the Spanish idiosyncrasy that solves the problem maintaining the advantages of the current system. The new system requires new constituencies to be drawn.
This paper states the redistricting problem and details the algorithm followed to solve it given the constraints that impose the Spanish multilevel governance system and the electoral system proposed. The algorithm avoids gerrymandering. An analysis of the solutions reached and of its consequences concludes the paper.

Keywords: Gerrymandering, multilevel governance, CCAA borders, drawing constituency boundaries.

Stein's Method, Fixed Point Biasing, Preferential Attachment Graphs and Quantum Mechanics

Erol Peköz
Boston University, USA

Stein's method is used to get error bounds for distributional limit theorems in difficult settings with dependence. One variant of the method uses a distributional fixed point equation for the limit to create biased random variables that couple closely and give Kolmogorov error bounds. The preferential attachment random graph model is used in network science to model growth of networks where heavily connected nodes

attract more future connections; this is a setting where there has been recent progress using this variant of Stein's method. I will survey some of this progress and give some new results in the multivariate approximation setting.

On Boundary Crossing by Stochastic Processes

Victor de la Peña
Dept of Statistics, Columbia University, USA

In this talk we introduce an approach to (sharply) bound the expected time for stochastic to cross a boundary. The approach can be thought as a direct extension of the concept of boundary crossing of non-random functions to that of stochastic processes. It can also be viewed as an extension of Wald's equations in sequential analysis to the case of stochastic processes with arbitrary dependence structure.

Partisan Bias in Multimember Districts

Alberto Penadés
Elections and Public Opinion Research Group (GIPEyOP), Universidad de Salamanca, Spain

Partisan bias is said to exist when two parties with equal number of votes are apportioned different number of seats. It is thus a dyadic relation not to be confounded with deviation for any specific standard (e.g. large party positive bias, the typical deviation from proportionality). Partisan bias has been aptly studied for electoral systems with single member districts, whereas for multimember systems the literature has insisted on proportionality indices. We extend the idea of partisan bias to any kind of (single tier) electoral system. Bias is measured as the additive result of an efficiency component (related to the heterogeneity of districts), a vote-weight component (related mainly to malapportionment and turnout rates) and third party effects.

Keywords: Partisan bias, Proportionality, Representation thresholds.

Stochastic Forecast Model of Severe Storm Wind over Territory of Northern Europe and England

Elvira Perekhodtseva
Moscow Technological University, Russia

The development of successful method for automated statistical well-in-advance forecast (from 12 hours to two days) of the summer severe storm winds, squalls and tornadoes is actual and very difficult problem in modern synoptic practice.
Nowadays in Russia there is no successful hydrodynamic model for the forecast of such storm wind (with the velocity V>24m/s), hence the main tools for the objective forecast development are the methods using the statistical model of these phenomena recognition. The values of the prognostic fields of some hydrodynamic Russian hemispheric and regional models, used in our statistic discriminant functions F(X), enabled us to develop the model of the automated hydrodynamic-statistical forecast of these storm winds over the territories of Europe and Russia including Siberia. These forecasts (up to 12-36hours ahead) for the territory of European part of Russia were recommended for the operative practice of Hydrometcenter of Russia. They are calculated automatically two times a day for the all territory of Europe. The verification of these storm wind forecasts for the different areas of Europe has shown their effectiveness in the Central and Northern Europe. Our successful forecast of St. Iuda storm over all the Northern Europe in October 2014 is a very interesting example. The other examples for this territory will be presented in the talk along with the maps of our automatic stochastic storm wind forecast. The stochastic forecast model of these phenomena (up to 60-72h ahead) on the base of the prognostic fields of the semi-Lagrangian model of the middle-term forecast will be developed next year.

Keywords: Storm Wind, Forecast, Stochastic Model, Hydrodynamic.

Outlier Detection and Identification when the Number of Outliers is Unknown

Lina Petkevičius, Vilijandas Bagdonavičius
Faculty of Mathematics and Informatics, Vilnius University, Lithuania

Outlier detection has been an important problem in many industrial and financial applications. We propose several new tests for outlier and contaminant detection and identification when the number of outliers is unknown. Classification power of these tests was investigated. A novel approach when robust estimators are used were proposed. Critical

values of the new and some known tests were found by simulation. Asymptotic values of these values were also found. The proposed tests have very good power with respect to well-known outlier detection tests. Real life data sets were investigated, too.

Keywords: Contaminants, Rosner's test, Outliers detection and identification, outlier tests.

Behaviour of Multivariate Tail Dependence Coefficients

Gaida Pettere[1], Ilze Zariņa[2], Irina Voronova[3]
[1]Dept of Engineering Mathematics, Riga Technical University, Latvia, [2]Actuarial Dept, AAS «BTA Baltic Insurance Company», Latvia, [3]Dept of Innovation and Business Management, Riga Technical University, Latvia

In applications an important property of a copula is tail dependence. Tail dependence shows the degree of dependence between random variables in the tail area. Bivariate tail dependence is investigated in many papers because of its importance in applications. Multivariate tail dependence has not been studied widely. The aim of this paper is to give a measure of the tail dependence for *n*-dimensional copulas. We have defined multivariate upper and lower tail dependence coefficients as limits of probability that values of one marginal will be large if at least one of other marginal will be as large too. Further we have tried to find some relations between introduced tail dependence and bivariate tail dependence coefficients. Gaussian copula doesn't have the tail dependence and therefore is not suitable for many applications. From another side applications have shown that the multivariate *t*-copula and skew *t*-copula can successfully be used in practice because of their tail dependence. Therefore we have paid our attention to the properties of introduced tail dependence coefficient for *t*-copula and skew *t*-copula using simulation technique.

Keywords: multivariate tail dependence, skew *t*-copula, *t*-copula.

Projecting Age-Specific Death Probabilities at Advanced Ages Using the Mortality Laws of Gompertz and Wittstein

Peter Pflaumer
Dept of Statistics, TU Dortmund, Germany

In this paper, death probabilities derived from the Gompertz as well as the Wittstein model are used to project mortality at advanced ages beginning at the age of 101 years. Life table data of Germany from 1871

to 2012 serve as a basis for the empirical analysis. Projections of the death probabilities and life table survivors will be shown. The proposed models are an alternative to the currently widely used logistic functions to fit observed probabilities at the oldest ages. In order to find the best model, it is necessary to have more data. Thus, the solution will be in the future, when the number of persons at advanced ages will significantly have increased.

Keywords: Mortality, Life Table, Mortality Deceleration, Centenarians.

On Fractional Stochastic Modeling for Biological Dynamics

Enrica Pirozzi
Dipartimento di Matematica e Applicazioni,
Università di Napoli FEDERICO II, Italy

The fractional calculus has ancient origins, although its potential application to the modeling of real phenomena is having only recently wide interest. Stochastic models based on fractional differential equations seem to be more suitable to provide descriptions and statistical predictions of complex biological phenomena, such as, for instance, the firing activity of neurons responsible for the transmission of information in cognitive processes [Teka et al., PLoS Comput Biol. 10(3): e1003526 (2014)], and the interaction between proteins able to develop work and movement, and for all dynamics generated by overlapping effects evolving over different scales of time and space.
Recently, in the context of the neuronal models, no time-homogeneous Gauss-Markov (GM) processes were used to describe the effect of a time-dependent input signal on the firing of the neuron [Buonocore et al., J. Comp Appl Math. Vol. 285. Pag.59–71 (2015)]; moreover, coupled GM processes and their modifications allowed to construct models for networks of neurons. The presence of particular boundaries in the state-space of the processes was useful to specialize the models: see the two-boundary model for the acto-myosin dynamics [D'Onofrio and Pirozzi, J. Math. Biol. (2016) doi:10.1007/s00285-016-1061-x]. In all above dynamics, different time scales are involved, such as those of the neuronal membrane voltage, ionic channels, myosin lever-arm, ATP-phase. The aim is that to consider the fractional stochastic models and investigate for an adequate description of the different time scales in such a framework. Starting from the fractional Brownian motion, fractional stochastic processes will be considered to extend the above stochastic models based on classical GM processes. Well-known neuronal models such as those including colored noise [Kobayashi et al., Fr in Comp Neurosci, 3-9 (2009)] will be revisited and generalized.

Keywords: Fractional derivatives, LIF neuronal model, Correlation.

AR-DENFIS for Mortality Data

Gabriella Piscopo
University of Genoa, Italy

In this paper we apply an integrated autoregressive dynamic evolving neuro-fuzzy inference system (AR-DENFIS) in the context of mortality projections. DENFIS is an adaptive intelligent system suitable for dynamic time series prediction, where the learning process is driven by An Evolving Cluster Method (ECM). The typical fuzzy rules of the neuro-fuzzy systems are updated during the learning process and adjusted according to the features of the data. This makes possible to capture the historical changes in the mortality evolution.

Keywords: DENFIS, ECM, mortality projections.

Contributions of Gilbert Saporta to Functional Data Analysis

Cristian Preda
Univ. Lille/ Inria Lille, France

We present the contributions of Gilbert Saporta to functional data analysis from his pioneer works (Méthodes exploratoires d'analyse de données temporelles, 1981). Functional PCA and canonical analysis for two functional random variables, functional regression methods and categorical functional data analysis will be presented.

Keywords: functional data.

References:

[1]. Saporta G. (1981) M'ethodes exploratoires d'analyse de données temporelles, Cahiers du B.U.R.O., No. 37-38, Université Pierre et Marie Curie, Paris.

[2]. C Preda, G Saporta, Clusterwise PLS regression on a stochastic process Computational Statistics & Data Analysis 49 (1), 99-108 (2005)

[3]. AM Aguilera, M Escabias, C Preda, G Saporta, Using basis expansions for estimating functional PLS regression: Applications with chemometric data Chemometrics and Intelligent Laboratory Systems 104 (2), 289-305 (2010)

The Compound Run Length Distribution: Properties and Applications

Athanasios C. Rakitzis[1], Markos V. Koutras[2]
[1]*Dept of Mathematics, University of Aegean, Greece,*
[2]*Dept of Statistics and Insurance Science, University of Piraeus, Greece*

In the present work we study a compound run length distribution of a control chart, by exploiting the fact that the distribution of the run length L is a discrete Phase-type one. The term compound run length distribution refers to the distribution of the compound random variable $S_L = \sum_{t=1}^{L} Y_t$, where $Y_1, Y_2, \cdots$, is a sequence of independent and identically distributed, positive valued random variables, independent of L. We illustrate how the performance of various control charts that are suitable for monitoring Poisson observations, can be evaluated in terms of the distribution of S_L. The suggested framework, provides a more realistic scenario, as compared to the classical control chart setup. An illustrative example is presented as well.

Keywords: Control charts, CUSUM, geometric distribution, runs rules, phase-type distributions, Poisson distribution, time between inspections, statistical quality control.

Acknowledgment: *M.V. Koutras has been partially funded by National Matching Funds 2014-2016 of the Greek Government, and more specifically by the General Secretariat for Research and Technology (GSRT), related to EU project "IS MPH: Inference for a Semi-Markov Process" (GA No 329128).*

Assessment of Clustering of Deaths among Families with Declining Levels of Mortality in India, 1992-2006

Mukesh Ranjan, L. K Dwivedi
International Institute for Population Sciences (IIPS), India

In the changing socio-economic environment in the country in the post liberalization period, the mortality levels have been declined substantially but we found that the pace of reduction in mortality is much faster than the pace in reduction of clustered deaths in families. Though the high risk families have declined but now almost similar level of clustered death is experienced by lower number of families. Utilizing the pooled retrospective birth history data of the three rounds of National Family Health Survey data (1992-2006) in random effect logit model, we found that after adjusting the socio-bio demographic factors in Model 2, the odds of infant deaths for interaction of time with previous death in the

family has increased but the Infant mortality has declined substantially as captured by the time factor and constant. Nearly 10 percent variation (intraclass correlation) in infant mortality is explained by the mother level unobserved factors.

Keywords: death clustering, national family health survey, families.

Tribal Death Clustering in Central and Eastern Indian States

Mukesh Ranjan, L. K Dwivedi
International Institute for Population Sciences (IIPS), India

Present study attempts to assess family level death clustering among mothers in the tribes of central and eastern Indian states of Jharkhand, Madhya Pradesh, Orissa and Chhattisgarh. Mixed effect model with random intercept by taking micro-data of National Family Health Survey-3 (2005-06), India was used for analysis. The raw data clustering analysis showed existence of clustering in all the four states with maximum clustering is among tribes of Chhattisgarh and least is in Jharkhand. The most important factor which increases the risk of infant deaths is the causal effect of infant death on the risk of infant death of the subsequent sibling (a scarring effect), after controlling for mother-level heterogeneity. The estimates reject the null hypothesis of no mother-level unobservable among the tribes of Jharkhand and Madhya Pradesh at conventional level of significance but, in the states of Orissa and Chhattisgarh unobservable at mother level have limited power to explain death clustering. Results also shows that high-risk tribal families are more exposed to short birth intervals and are likely to reach higher parities in their attempts to achieve their desired family size.

Keywords: Death clustering, Central and Eastern India, Unobserved heterogeneity, family.

A Biparametric Version of the Univariate Generalized Waring Distribution

Jose Rodríguez-Avi, Maria José Olmo-Jiménez, Valentina Cueva-Lopez
Dept of Statistics and Operational Research, University of Jaen, Spain

The Univariate Generalized Waring (UGW) distribution, with parameters (a, k, rho) is a triparametric distribution with non-negative parameters. It may be obtained as a two-step mixture from a Poisson distribution and

whose probability mass function is expressed in terms of the Gaussian function 2F1. It has a wide set of properties, such as the partition of the variance into three components: randomness, proneness and liability. In addition, a count data regression model based on this distribution has been developed.
Nevertheless, when we estimate its parameters by the method of maximum likelihood, the estimates obtained for the parameters a and k are similar, almost equal, many times. In this work, we present a biparametric distribution that may be included as a particular case of the UGW distribution when the two first parameters are equal. We show the main probabilistic properties of this distribution. Moreover, we compare it with the UGW as well as with other overdispersed biparametric distributions and we make a simulation study in relation to estimation where we show that, in many cases, results provided by this model in a UGW scenario are quite similar, and even better in terms of goodness of fit and Akaike Information Criteria, than those provided by the triparametric model. Finally, some application examples to real data are included.

Keywords: Count data distribution, Univariate Generalized Waring distribution, Goodness of fit.

Schur Properties of Convolutions of Gamma Random Variables with Applications in RandNLA

Farbod Roosta-Khorasani
International Computer Science Institute and Dept of Statistics, University of California at Berkeley, USA

An important problem in Randomized Numerical Linear Algebra (RandNLA) is estimating the trace of an implicitly given matrix, which arises in many scientific and data analysis applications. One such estimator is built using Gaussian random vectors. The analysis of this estimator gives rise to questions regarding the stochastic ordering among convolutions of heterogeneous gamma random variables. Sufficient conditions are discussed for comparing such convolutions in terms of the usual stochastic order. These comparisons are characterized by the Schur convexity properties of the cumulative distribution function of the convolutions.

Analysis of the Determinants and Outputs of Innovation in the Nordic Countries

Cátia Rosário[1], António Augusto Costa[2], Ana Lorga da Silva[3]
[1]Centro de Pesquisa e Estudos Sociais, Escola de Ciências Económicas e das Organiza-ções; Universidade Lusófona de Humanidades e Tecnologias, Portugal, [2]Escola de Ciências Económicas e das Organizações; Universidade Lusófona de Humani-dades e Tecnologias, Portugal, [3]Centro de Pesquisa e Estudos Sociais, ECEO and FCSEA, Universidade Lusófona de Humanidades e Tecnologias, Portugal

This study was applied to the European Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) which are referred by the European Commission as countries with high innovation performance.
The analysed panel data concern the period between 1999 and 2014 and it was studied how different inputs of innovation affect the different outputs.
The "innovation" variable was constructed using factor analysis, given that the Organisation for Economic Co-operation and Development considers that innovation is the result of a set of macro measures common to different countries. The factor obtained through the exploratory factor analysis represents the results of innovative activity and economic performance.
It was analysed how the quality of human capital, the research and development efforts carried out by different economic agents affect the results of innovation. It was possible to conclude that countries with a higher proportion of applied research and more cooperation between researches carried out by companies, universities and the government lead to better economic results and to higher outcomes of intellectual property.

Keywords: Innovation, Panel Data, Regression Models, Factorial analysis.

Non-Metric Partial Least Squares for Non-linear Structural Equation Model estimation

Giorgio Russolillo[1], Laura Trinchera[2]
[1]Conservatoire national des arts et métiers, France,
[2]NEOMA Business School, France

Partial Least Squares approach to latent variable path modeling (PLS-PM) is a component-based method that allows investigating the relationships among several sets of variables connected by a path of relations. This method is implemented through a flexible algorithm that, depending on the chosen option, can optimize covariance or correlation

based criteria involving components obtained as linear combinations of the variables in each set.
PLS-PM has been originally proposed as a component-based alternative to factor-based Structural Equation Modeling (SEM). The use of PLS-PM for estimating parameters of a factor-based SEM has been often criticized, as it yields biased estimates for model parameters. However, recent literature has reconsidered PLS-PM algorithm as an estimation procedure for factor-based SEM, as a new procedure, named consistent PLS (PLSc), has been shown to yield consistent and asymptotically normal estimators for parameters of linear SEM.
The Non-Metric approach to PLS-PM (NM-PLSPM) has been recently introduced in order to properly include in PLS-PM non-metric variables. NM-PLSPM is a modified PLS-PM algorithm that works as an optimal scaling tool. In NM-PLSPM non-metric variables are quantified so as to optimize PLS-PM criteria under two set of parameters: the PLS parameters (the weights for obtaining the components) and the scaling parameters (a set of numerical values that replaces the levels of the non-metric variables).
In this work we propose to merge NM-PLSPM and PLSc in order to consistently estimate non-linear SEM.

Keywords: SEM, PLS-PM, non-linearity, NM-PLSPM.

Reliability of Decrease Rates for Cardiovascular Mortality in Russia

Tamara P. Sabgayda, Victorya G. Semyonova
Dept of analysis of health statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Russia

The reduction rates of cardiovascular mortality (which observed since 2003) increased sharply in recent years. Paradoxically, the timing of this change coincided with the timing of new order providing medical care to patients with cardiovascular disease, which is under increased government attention. *The aim*: to test the consistency changes mortality from non-communicable diseases in adult population.
Unnaturally sharp increase in mortality from 2013 was found for diabetes mellitus (twice), nervous system diseases (2.2 times for men and 3.0 times for women), psychiatric disorders (1.9 and 3.3-fold), diseases of genitourinary system and skin (one third), diseases of musculoskeletal system (1.8 and 2.0 times). For all classes of death causes the greatest growth has occurred in the oldest age group (75+). Only for cardiovascular disease the average age of deceased decreased from 2013 to 2015 and it has increased for all other death causes classes.

Predicted and observed death rates among total and elderly populations differ on average by 15.0% and 19.8% respectively for cardiovascular disease, for diabetes by 57.0% and 61.6%, for nervous system diseases by 54.5% and 75.7%, for mental disorders by 63.8% and 82.5%, for diseases of genitourinary system by 30.1% and 36.0%, for skin diseases by 20.7% in both cases, for diseases of musculoskeletal system by 42.5% and 63.0%.

Thus, deliberate choice of avoiding cardiovascular disease as the underlying death cause for elderly persons has caused an extremely high rate of decline in mortality from circulatory diseases. Such situation does not mean underestimation of cardiovascular mortality as namely these death causes were selected by default for coding undiagnosed causes of death of elder people previously.

Keywords: cardiovascular mortality, predicted and observed death rates, average age of deceased, correction of mortality structure.

Results on Multivariate Risk-Adjusted Survival Time CUSUM and EWMA Control Charts

Athanasios Sachlas[1,2], Stelios Psarakis[1], Sotiris Bersimis[2]
[1]Dept of Statistics, Athens University of Economics and Business, Greece,
[2]Dept of Statistics and Insurance Science, University of Piraeus, Greece

In the last two decades, a modification of standard and advanced control charts appeared in the bibliography to improve the monitoring mainly of medical processes. This is the risk-adjusted control charts, which take into consideration the varying health conditions of the patients. Biswas and Kalbfleisch (2008) outlined a risk-adjusted CUSUM procedure based on the Cox model for a failure time outcome while Sego et al. (2009) proposed a risk-adjusted survival time CUSUM chart for monitoring a continuous, time-to-event variable that may be right-censored. Motivated by the above mentioned papers, in this work we present a review of the literature on risk-adjusted charts and we provide some preliminary results on multivariate risk-adjusted survival time CUSUM and EWMA control charts.

Keywords: Risk-adjusted control charts, multivariate control charts.

References

[1]. P. Biswas and J.D. Kalbfleisch. A risk-adjusted cusum in continuous time based on the cox model. Statistics in Medicine, 27:3382-3406, 2008.

[2]. L.H. Sego, M.R. Reynolds Jr, and W.H. Woodall. Risk-adjusted monitoring of survival times. Statistics in Medicine, 28:1386-1401, 2009.

News Augmented GARCH (1,1) Model for Volatility Prediction

Zryan Sadik, Paresh Date
Brunel University London, United Kingdom

Forecasting of stock return volatility plays an important role in the financial markets. Applying a GARCH models to stock return time series is one of the established methods for predicting volatility. In this study, we have considered quantified news sentiment as a second source of information, which is used together with the market data to predict the volatility of asset price returns. We call this news augmented GARCH model NA-GARCH. Our empirical investigation compares volatility prediction of returns of 12 different stocks (from two different stock markets), with 9 data sets for each stock. Our results clearly demonstrate that NA-GARCH provides a superior prediction of volatility than the plain vanilla GARCH model. Our findings also show that the positive news tends to significantly reduce volatility whereas negative news tends to increase volatility. These results vindicate some recent findings regarding the utility of news sentiment as a predictor of volatility, and also vindicate the utility of our novel model structure combining the proxies for past news sentiments and the past asset price returns. NA-GARCH is thus a computationally efficient means of exploiting the news sentiment score for better volatility or VaR prediction and it has a potential to be very useful in industrial practice.

Keywords: Volatility prediction, GARCH, NA-GARCH, news sentiment, news impact scores.

Defective Galton-Watson Processes

Serik Sagitov[1], Carmen Minuesa[2]
[1]Dept of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Sweden,
[2]Dept of Mathematics, University of Extremadura, Spain

The Galton-Watson process is a Markov chain modelling populations in which each individual reproduces independently of the others giving birth to k offspring with probability p_k, $k \geq 0$.
In this work we study defective Galton-Watson processes having defective reproduction laws, that is, $\sum_{k\geq 0} p_k = 1 - \varepsilon$, for some $\varepsilon \in (0, 1)$. In this setting, each particle may send the process to a graveyard state Δ with probability ε. These processes have state space $\{0, 1, \ldots\} \cup \{\Delta\}$ and become eventually absorbed either at 0 or at Δ. Since in realistic settings, the defect ε of the reproduction is small, we analyse the

trajectories of such processes as $t \to \infty$ and $\varepsilon \to 0$, assuming that the process has avoided absorption until the observation time t.

Keywords: branching process; defective distribution; Galton-Watson process with killing.

Robust Bayesian Analysis using Classes of Priors

Sánchez-Sánchez M., Sordo M. A., Suárez-Llorens A.
Departamento de Estadística e Investigación Operativa,
Facultad de Ciencias, Universidad de Cádiz, España

In the context of robust Bayesian analysis, we focus on a new class of prior distributions based on stochastic orders and distortion functions defined in Arias-Nicolás et al. (2016). We will apply this new class in different contexts. Namely, we will analyse the problem of computing different premium principles in risk theory. We will consider that uncertainty with regard to the prior distribution can be represented by the assumption that the unknown prior distribution belongs to the new class of distributions and we will examine the ranges of the Bayesian premium when the priors belong to such a class. Kolmogorov and Kantoverich metrics could be a good election to measure the uncertainty induced by such class, as well as its effect on the Bayesian Premiums. Finally, we will also discuss the extension to the multivariate case. We will provide new definitions and their interpretations.

Keywords: Robust Bayesian Analysis, prior class, stochastic orders, distortion functions, premiums.

Reference: Arias-Nicolás, J.P., Ruggeri, F. and Suárez-Llorens, A. New classes of priors based on stochastic orders and distortion functions. Bayesian Analysis, 11, 4, pp. 1107-1136, 2016.

A Realized p-Variation Random Function as a Statistical Diagnostic for Semimartingales

Lino Sant
Dept of Statistics and Operations Research,
Faculty of Science, University of Malta, Malta

The stochastic properties of the p-variation of semimatingales, and more recently the statistical properties of realized p-variation, have played important and crucial roles in the study and applications of semimartingales respectively. Realized p-variation can help solve problems involving hypothesis testing, diagnostic checking and parameter estimation for stochastic processes. In fact within the context

of Levy processes the independent, identically distributed increments property largely determines the statistical behaviour of these sample path derived statistics. Realized variation can be split into parts originating from the canonical components of semimartingales for which many results have been proved especially in the general setting of Ito semimartingales.
In this paper a random function is defined out of sample path readings with the power p serving as argument: p→∑_(i=1)^n▒|X_(t_(i+1))-X_(t_i) |^p . Its properties and performance are studied eventually within a Banach space context wherein issues of stochastic equicontinuity ensure uniform convergence. This random function is being proposed as a general purpose method for investigating the nature of the generating process. Functionals derived from it could also be very revealing in the course of identifying the type of process one is dealing with and obtaining estimates of the relevant parameters. Statistical results and simulation runs are proposed and discussed.

Keywords: Ito semimartingales, realized p-variation, random function.

Choosing Tuning Instruments for Generalized Rubin-Tucker Lévy Measure Estimators

Lino Sant, Mark Anthony Caruana
Department of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

Estimation of the Lévy measure through the increments of a Lévy process is a problem which has attracted much attention over the last decade. The first such estimator comes from Rubin and Tucker but it has not been much in use. Its performance is not satisfactory and most researchers have tried alternatives. Various issues surround this statistical problem, most notably the behavior of the Lévy measure at the origin. Even increments from BM yield poor estimates. But Rubin-Tucker type estimators, that is distribution function estimators constructed out of increments $X_{i,j}$ coming from sample path values proposed in the more general form:

$$\hat{H}(u) = \frac{1}{N}\sum_{i=1}^{N} n^{b} \sum_{j=1}^{n} \frac{|X_{i,j}|^{\varphi(X_{i,j})}}{1+|X_{i,j}|^{\varphi(X_{i,j})}} 1_{\{|X_{i,j}|<u}$$

can be improved.
In this paper the authors study this estimator and look for suitable choices of the tuning parameter *b* and the function φ to improve estimator quality and convergence rates. Various classical and other recent results are put into use to obtain an estimator for an equivalent Lévy measure distribution function suitably transformed. The choice of the function φ

was guided by convergence results, in particular, the choice $\varphi(x) = 2 - x^{\beta}$ for $-1 \le x \le 1$, with β close to 0 on the positive side, is shown to have special benefits which are studied in this paper.

Keywords: Lévy measure, Lévy process, convergence rates, distribution function.

An Approximation to Social Wellbeing Evaluation using Structural Equation Modeling

Leonel Santos-Barrios[1], Mónica Ruiz-Torres[2], William Gómez-Demetrio[1], Ernesto Sánchez-Vera[1], Ana Lorga da Silva[3], Francisco Martínez-Castañeda[1]
[1]ICAR-Universidad Autónoma de Estado de México, México,
[2]IIZD-Universidad Autónoma de San Luis Potosí, México,
[3]ECEO; CPES- Universidade Lusófona de Humanidades e Tecnologias, México

In order to evaluate how small-scale livestock models, contribute to the social wellbeing "status", we use structural equation modeling. Five latent variables were included in the model according to Keyes theory (Keyes, 1998) of that represent the extent to which individuals are overcoming social challenges and are functioning well in their social world. The variables were: social integration; social acceptance; social Contribution; social actualization; and social coherence.

Keywords: Social wellbeing, Structural Equation Modelling, Psychological Behaviour, Livestock.

Clinical Trials Simulation: Comparison of Discrete Method, Continuous Method and Copula Method for virtual Patients' Generation

Nicolas Savy[1], Philippe Saint-Pierre[1], Sébastien Déjean[1], Stéphanie Savy[2], Sébastien Marque[3]
[1]Toulouse Institute of Mathematics, France,
[2]Estrials, France, [3]Capionis, France

Clinical trials is a complex and challenging process essentially for ethical, financial and scientific concerns. For twenty years, simulated clinical trials (SCT) have been introduced in the drug development. It has become more and more popular mainly due to pharmaceutical companies which aim to optimize their clinical trials (duration and expenses) and to the regulatory agencies which consider simulations as an alternative tool to reduce safety issues and fasten evaluation process. The whole simulation plan is based on virtual patients' generation which

consists in the random generation of vectors of covariates describing the baseline informations of a sample of (virtual) patients. To be relevant, the structure of the sample must be as close as possible to what is actually observed (marginal distributions and correlation structure).
The simplest and easiest way, referred as Discrete methods, could be to perform Monte Carlo simulations from the joint distribution of the covariates. This is trivial when the parameters of the distribution are known, but, on concrete examples, available data may come from historical databases, which imply to have a preliminary estimation step. For Discrete method this step may not be effective especially when there are a lot of covariates mixing continuous and categorical ones. In this paper, simulation studies illustrates that two alternative methods (the Continuous method and the copula method) may be good alternative to the Discrete one especially when marginal distributions are moderately bimodal.

Keywords: Copula, Monte Carlo, Virtual Patients' Generation, Simulated Clinical Trials.

Stochastic Correlation in Energy Markets

Luis Seco
University of Toronto, Canada

Energy markets work under a time structure where backwardation and contago variations are the source of primary risks; these are described via multifactor models which, if assumed Gaussian, will rarely produce variations as observed in the market. In this talk we will review non gaussian models based on stochastic correlation which produce results more in line with observations.

Keywords: Correlation, stochastic correlation, regime switching models, mathematical finance, risk management.

Migration Component in Health Losses of Population in Russian Megapolis (for example of Moscow)

Victorya G. Semyonova[1], Tamara P. Sabgayda[1], Svetlana Yu. Nikitina[2]

[1]Dept of analysis of health statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Russia, [2]Dept of Statistics and Population Health of the Federal State Statistics Service; Russia

Migration processes are increasingly becoming a factor influencing the formation of modern societies throughout the world including Russia, which is the 2nd country on migration attractiveness after the United States. This put the acutely actual issue: to what extent migrants affect the losses of Russian public health? At the moment the losses of Moscow's population are determined by undocumented migrants. They constitute half in children under one year old, up to 40% in children under 18, among working population more than a quarter in men and 20% in women. Only in death number of older persons the proportion of undocumented migrants becomes imperceptible (about 6%). The high migration component of infant, child and adolescent mortality is mainly determined by external causes. The greatest contribution of undocumented migrants of working age in Moscow mortality observed for exogenous causes (injuries and poisoning, infectious diseases, ill-defined status, diseases of the digestive system). The risks of death from the vast majority of causes in all major age groups among persons who are not registered in Moscow are much higher than those for the permanent population of the megapolis.

In 2014, for people who were registered in Moscow, life expectancy amounted to 76.3 years for male and 82.3 years for female while for total Moscow population they were 73.2 years and 80.8 years respectively. That is, Moscow is losing more than 3 years of male life expectancy and 1.5 years of female life expectancy due to undocumented migrants.

Thus, undocumented immigrants are the primary group of premature mortality risk, on the one hand, and massive reserve of life expectancy growth in the megapolis, on the other hand. The effective improvement of megalopolis residents' health is possible in the first place by addressing the problems of persons who are not registered there.

Keywords: undocumented immigrants, life expectancy growth, population of megapolis, mortality.

L-Comoments: Theory and Applications

Robert Serfling
Dept of Mathematical Sciences, University of Texas at Dallas, USA

For measuring spread of a univariate distribution, a now classical alternative to the standard deviation is the Gini mean difference (Gini, 1912), which is defined under merely first moment assumptions and is less sensitive to extreme observations. Likewise, the "Gini covariance" (Schechtman and Yitzhaki, 1987) is an analogue of the usual covariance but requires only first moments. The "L-moments" (Hosking, 1990) provide an entire series of univariate descriptive measures (location, dispersion, skewness, kurtosis, etc.), the first order case being the mean and the second order case the Gini mean difference, all available under just first order assumptions. These are alternatives to the usual central moments. For multivariate distributions, the usual covariance matrix requires second moments and the higher order "central comoments" (Rubinstein, 1973) require increasingly higher order moments. Alternatively, however, the "L-comoments" (Serfling and Xiao, 2007) extend L-moments to the multivariate setting yet require only first moments regardless of order of comoment. In the time series setting, the Gini covariance recently has become applied to formulate a "Gini autocovariance function"(Serfling, 2010, Shelef and Schechtman, 2011, Carcea and Serfling, 2015) available under merely first order assumptions. This talk provides an overview of these various developments, mentions several application contexts, and indicates recent related work.

Keywords: Multivariate, Nonparametric, Comoments, Time series, Autocovariance.

A Gini-Based Time Series Analysis and Test for Reversibility

Amit Shelef[1], Edna Schechtman[2]
[1]Dept of Logistics, Sapir Academic College, Israel,
[2]Dept of Industrial Engineering and Management,
Ben-Gurion University of the Negev, Israel

Time reversibility is a fundamental hypothesis in time series. In this research, Gini-based equivalents for time series concepts that enable to construct a Gini-based test for time reversibility under merely first-order moment assumptions are developed. The key idea is that the relationship between two variables using Gini (as measured by Gini autocorrelations and partial autocorrelations) can be measured in two directions, which are not necessarily equal. This implies a built-in

capability to discriminate between looking at forward and backward directions in time series. The Gini creates two bi-directional Gini autocorrelations (and partial autocorrelations), looking forward and backward in time, which are not necessarily equal. The difference between them may assist in identifying models with underlying heavy-tailed and non-normal innovations. Gini-based test and Gini-based correlograms, which serve as visual tools to examine departures from the symmetry assumption, are constructed. Simulations are used to illustrate the suggested Gini-based framework and to validate the statistical test. An application to a real data set is presented.

Keywords: Autocorrelation; Autoregression; Gini correlation; Gini regression; Moving block bootstrap; Time reversibility.

Some Extensions in Space-Time LGCP: Application to Earthquake Data

Marianna Siino[1], Giada Adelfio[1], Jorge Mateu[2]
[1]*Dip. di Scienze Economiche Aziendali e Statistiche, University of Palermo, Italy,*
[2]*University Jaume I, Spain*

In this paper we aim at studying some extensions of complex space-time models, useful for the description of earthquake data. In particular we want to focus on the Log-Gaussian Cox Process (LGCP, [1]) model estimation approach, with some results on global informal diagnostics. Indeed, in our opinion the use of Cox processes that are natural models for point process phenomena that are environmentally driven could be a new approach for the description of seismic events. These models can be useful in estimating the intensity surface of a spatio-temporal point process, in constructing spatially continuous maps of earthquake risk from spatially discrete data, and in real-time seismic activity surveillance. Moreover, covariate information varying in space-time can be considered into the LGCP model, providing complex models useful for a proper description of seismic events. LGCP is a Cox process with $\Lambda = expS(x)$, where S is a Gaussian process. This construction has some advantages, related to the multivariate Normal distribution features, since the moment properties $\Lambda(x)$ are inherited by the Cox process. In particular, both estimation and diagnostics, can deal with some higherorder properties [2], expressed for instance by the intensity and the pair correlation function of the LGCP.

Keywords: LGCP, Space-time Point Processes, second-order functions, diagnostics.

References:

[1]. Moller J, Syversveen AR, Waagepetersen RP. Log Gaussian Cox Processes. Scandinavian Journal of Statistics, 25(3), 451-482. 1998.

[2]. Baddeley AJ, Moller J, Waagepetersen R. Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns. Statistica Neerlandica, 54, 329-350. 2000

Acknowledgements: *This paper has been supported by the national grant of the MIUR for the PRIN-2015 program, \Prot. 20157PRZC4 – Research Project Title Complex space-time modeling and functional analysis for probabilistic forecast of seismic events. PI: Giada Adelfio"*

Doubly Recurrent Algorithms for Mixed Power-Exponential Moments of Hitting Times for Semi-Markov Processes

Dmitrii Silvestrov[1], Raimondo Manca[2]
[1]Dept of Mathematics, Stockholm University, Sweden, [2]Dept of Methods and Models for Economics, Territory and Finance, University "La Sapienza" Rome, Italy

New doubly recurrent algorithms for computing mixed power-exponential exponential moments of hitting times and accumulated rewards of hitting type for semi-Markov processes are presented. The algorithms are based on special techniques of sequential phase space reduction and recurrence relations connecting mixed power-exponential exponential moments of hitting times. Applications are discussed as well as possible generalizations of presented results and examples.

Keywords: Semi-Markov process, Hitting time, Mixed power-exponential moment, Recurrent algorithm.

Asymptotic Recurrent Algorithms for Nonlinearly Perturbed Semi-Markov Processes

Sergei Silvestrov[1], Dmitrii Silvestrov[2]
[1]Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Sweden,
[2]Dept of Mathematics, Stockholm University, Sweden

Recurrent algorithms for construction of Laurent asymptotic expansions for power moments of hitting times for nonlinearly perturbed semi-Markov processes are presented. These algorithms are based on a special technique of sequential phase space reduction, which can be applied to processes with an arbitrary asymptotic communicative structure of phase spaces. Asymptotic expansions are given in two forms, without and with explicit upper bounds for remainders. A special attention is paid to nonlinearly perturbed birth-death-type semi-Markov processes. Applications to nonlinearly perturbed queuing systems,

information networks, and models of stochastic systems of biological nature are discussed.

Keywords: Semi-Markov process, Nonlinear perturbation, Hitting time, Power moment, Laurent asymptotic expansion.

Simulation Before, During, and After a Clinical Trial: A Bayesian Approach

Stephen D. Simon[1,2]

[1]*Dept of Biomedical and Health Informatics, University of Missouri-Kansas City, USA,* [2]*P.Mean Consulting, USA*

Simulation of a clinical trial gives you answers to important economic, logistical, or scientific questions about the trial when some of the features are difficult to characterize with perfect precision. A Bayesian approach with informative priors offers a flexible framework for trial simulation. It provides a seamless transition from simulation prior to the trial to simulation during the trial itself. Although informative priors are controversial, you can avoid perceptions of bias by restricting the informative priors to clinical trial features that are independent of your research hypothesis. You can protect your interim predictions against unrealistic prior beliefs by implementing the hedging hyperprior, a simple hyperdistribution that downweights the strength of the prior when there is a discrepancy between the prior distribution and data observed during the trial itself. The Bayesian approach also gives you a simple post mortem analysis after the trial ends. You can compute percentile values by plugging the point estimates from the actual clinical trial data into the corresponding prior distributions. Over multiple trials, a deviation in these percentiles from a uniform distribution indicates biased specification of the informative priors. The Bayesian approach to trial simulation will be illustrated using various patient accrual models.

Keywords: hedging hyperprior; informative prior distributions; Markov Chain Monte Carlo; patient accrual.

Estimation of the Healthy Life Expectancy in Italy through a Simple Model based on Mortality Rate

Christos Skiadas[1], Maria Felice Arezzo[2]
[1]*ManLab, Technical University of Crete, Greece,*
[2]*Università di Roma "La Sapienza", Italy*

We use an advanced methodology based on the mortality rate of a population to explore the healthy life expectancy (HALE) estimates of the World Health Organization (WHO) from the Global Burden of Disease Study. First we calculate the loss of healthy life year estimator (LHLY) and then the healthy life expectancy (HLE). We use the full life tables from the Human Mortality Database (HMD). Our estimates are compared with the HALE estimates for Italy and other countries. Another direct estimation based on the *mx* and *qx* quantities provided from the Life Tables is also introduced and tested.
Analysis of the health situation in Italy males and females is presented along with the healthy life expectancy estimates.

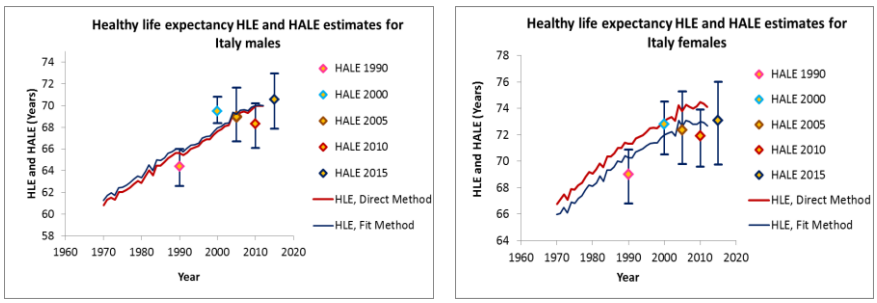


Fig. HLE estimates and HALE estimates and confidence intervals for Italian males (left) and females (right)

Keywords: Healthy life expectancy, World Health Organization, Global burden of disease study, Life tables, Healthy life expectancy in Italy, Healthy life years lost, WHO, HALE.

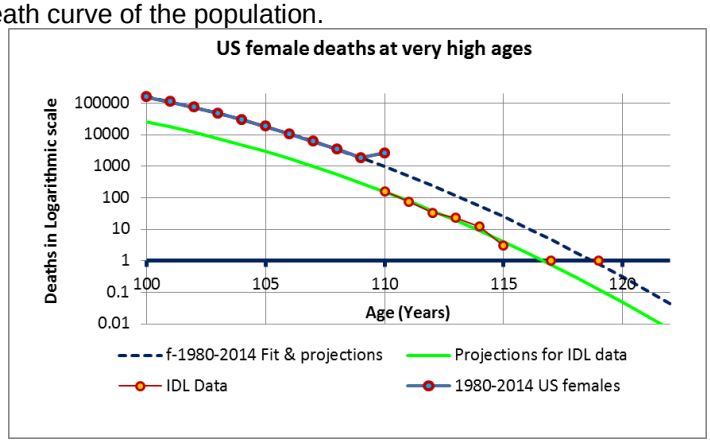
Further Exploration of the Existence of a Limit to Human Life Span

Christos H. Skiadas[1], Charilaos Skiadas[2]
[1]*ManLab, Technical University of Crete, Greece,*
[2]*Department of Mathematics and Computer Science, Hanover College, USA*

The probability of dying is a distribution with a special tail to the right. The tail is approaching zero for a high theoretical age. Theoretically the probability of dying at high ages could be very small and it is hard to

appear such large population size to achieve at least one survival at very high ages. The question addressed related to a limit of human life span is mainly influenced by the population needed to achieve a certain limit. For a population of infinite size the probability of at least one survival at a very high age is possible. However, for the limited global population an upper limit could exist at least of a probabilistic point of view. The decline of the probability of dying at high ages is very fast. It exponentially reduces the chances of survival to high ages so that, following our estimates, after estimating the age year for the last survival, achieving one more survival at the next year of age, we need approximately a ten times higher population.

We provide a data transformation method followed by fitting and the related projections are important tools to find the time course of the super-centenarians development. A model application method is also applied to cross-validate the previous estimates in the extreme right of the death curve of the population.



The fit and the projections are illustrated in the figure above in logarithmic scale. The female deaths provided from the Human mortality Database (HMD) for 110+ for 1980-2014 are 2576 whereas the projected are 1930 covering the 75% of the total sum. The provided 309 deaths from IDL data is the basis for finding the trajectory for this case presented by the green line in the same figure. We multiply by 0.16 the HMD projections. The used methodology provides a good method to locate the trajectory for the IDL data whereas it can estimate some extra points placed outside this trajectory by means the case surviving at 119 years of age. The latter case is within the space located by the upper curve which is produced from the 1980-2014 population size that is 39.617.539 and 2.576 deaths for 110+ years of age. The super centenarians' deaths from IDL data base refer to a population size of 6.338.806 that is close to a five year sum of death females in US.

We thus have a method to relax the debate rased after the publication of the paper in Nature by Xiao Dong, Brandon Milholland & Jan Vijg on "Evidence for a limit to human lifespan".

Keywords: Centenarians, super centenarians, HMD, Human life span limit, Fitting method, Projection method, Data transformation method.

References:

[1]. Skiadas, Charilaos and Skiadas, Christos, H. Development, Simulation and Application of First Exit Time Densities to Life Table Data, *Communications in Statistics* 39, 2010: 444-451.

[2]. Skiadas, Christos, H. and Skiadas, Charilaos. Exploring the State of a Stochastic System via Stochastic Simulations: An Interesting Inversion Problem and the Health State Function. *Methodology and Computing in Applied Probability* (2015, Volume 17, Issue 4, pp 973–982).

[3]. Xiao Dong, Brandon Milholland & Jan Vijg. Evidence for a limit to human lifespan, *Nature*, Vol 538, 2016:257-259.

Comparing Generalized Mixed Effects Models with RE-EM Tree Method in Corporate Financial Distress Prediction

Lukas Sobisek[1], Maria Stachova[2]

[1]Dept of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics Prague, Czech Republic, [2]Dept of Quantitative Methods and Informatic Systems, Faculty of Economics, Matej Bel University, Slovakia

An idea of our contribution is based on the previous studies that show a strong relationship between financial health and financial ratios of companies. Both cross-sectional or longitudinal financial data sets are analyzed in order to describe this relationship. We have chosen to investigate latter data set type collected over a several consecutive years. A longitudinal financial ratios database is used to predict financial health of companies.

To carry out our analysis we apply generalized mixed effects regression models in the statistical system R, namely the functions glmer() from lme4 package. The results are compared with those obtained by using RE-EM tree method, which combine the advantages of linear mixed models and classical CART algorithm by Leo Breiman. This method is implemented in REEMtree package in R.

Keywords: Generalized Mixed Effects Models, Financial Health of Companies, RE-EM tree.

Interpolation using Local Iterated Function Systems

Somogyi I., Soos A.
Faculty of Mathematics and Computer Science,
Babes-Bolyai University, Romania

Iterated function systems are used in order to construct fractal functions. Local iterated function systems are important generalization of iterated function systems. Also the real data interpolation methods can be generalized with fractal functions In this article using the fact that graphs of piecewise polynomial functions can be written as the fixed points of local iterated function system we study the behavior of local iterated function systems using interpolation methods.

Keywords: fractal function, iterated function system, local iterated function system, attractor.

A New Family of Premium Principles Obtained by a Risk-Adjusted Distribution

Miguel Ángel Sordo, Antonia Castaño, Gema Pigueiras
Dept of Statistics and Operation Research, University of Cádiz, Spain

Risk-adjusted distributions are commonly used in actuarial science to define premium principles. In this work, que claim that an appropriate risk-adjusted distribution, besides of satisfying other desirable properties, should be well- behaved under conditioning with respect to the original risk distribution. Based on a sequence of such risk-adjusted distributions, we introduce a family of premium principles that gradually incorporate the degree of risk-aversion of the insurer in the safety loading. Member of this family are particular distortion premium principles that can be represented as mixtures of TVaRs, where the weights in the mixture reflect the attitude toward risk of the insurer. We make a systematic study of this family of premium principles.

Keywords: premium principle, risk measure, order statistics, distortion function.

Selecting Speech Spectrograms to Evaluate Sounds in Development

Dimitrios Sotiropoulos
Technical University of Crete, Greece

In order to assess progress in developmental speech, it is necessary to evaluate the quality of vowel and consonant sounds in the production of words. This paper addresses the latter by proposing a novel method of selecting spectrograms from the large number that are available in speech data. An established method in the literature is to use the productions of each consonant in *all* the words in the data and by examining their spectrograms to decide to what proportion is the consonant produced accurately, that is, in an adult-like manner. The difficulty in achieving this lies in the fact that each word is produced more than once and is not produced in the same way every time, resulting in a large amount of computations. To overcome this, it is proposed here that a consonant is considered to be accurately produced in a specific word if it is accurately produced at least once. This way, the amount of computations is reduced since the remaining productions of this word are ignored, and the proportion of accurate productions of a consonant in a specific word is either one or zero. It is clear that when consonants are produced either always accurately or never accurately in a specific word, the two methods yield the same result. During speech development, however, children's productions of consonants vary. Therefore, the pertinent question is whether the proposed approximation of a consonant's accuracy is satisfactory as compared to the accuracy computed from the whole speech data, and how this comparison varies across all the consonants. This question is answered here by considering a child's English speech productions over one month during phonological development, at two and a half years old. The proposed approximate proportion of accuracy over all consonants has a mean of 0.46 and a variance of 0.12. If the exact proportion of accuracy is taken as equal to the approximate proportion of accuracy, the error involved results in a coefficient of determination equal to 0.87. If the exact proportion is obtained from the approximate proportion by raising it to the power of 1.67, the coefficient of determination is improved becoming 0.97. The results obtained here should encourage similar calculations for other children at different stages in their speech development with the aim of obtaining a universal formula for computing consonant accuracy using only a sample of speech spectrograms from the complete set in a speech data.

The Epistemic Uncertainty Analysis for the Inventory Management System with (Q,r) Policy

Massinissa Soufit, Karim Abbas
Bejaia University, Algeria

In this paper, we consider the inventory systems, under propagation of epistemic uncertainty in some model parameters. The applying of Taylor series expansion method for Markov chains, we compute performance measures for a class of inventory systems, under the epistemic uncertainty inflicted in the model parameters. Specially, we calculate the expected value and the variance of the stationary distribution associated with the considered model. Various numerical results are presented and compared to the analogously Monte Carlo simulations ones.

Keywords: Inventory; Parametric uncertainty; Taylor series expansions; Monte Carlo Simulation.

At-Risk-of-Poverty or Social Exclusion Rate – Regional Aspects in the Slovak and Czech Republic and International Comparison

Iveta Stankovičová[1], Jitka Bartošová[2], Vladislav Bína[2]
[1]Dept of Information Systems, Faculty of Management, Comenius University, Slovakia, [2]University of Economics Prague, Faculty of Management, Czech Republic

More than 120 million people are at risk of poverty or social exclusion in the EU. EU leaders have pledged to bring at least 20 million people out of poverty and social exclusion by 2020. The fight against poverty and social exclusion is at the heart of the Europe 2020 strategy for smart, sustainable and inclusive growth. Each individual member state will have to adopt one or several national targets.
Presented article examines the aggregate indicator of poverty and social exclusion AROPE in the Slovak and Czech Republic. Indicator AROPE is the sum of persons who are at-risk-of-poverty or severely materially deprived or living in households with very low work intensity as a share of the total population. Source for calculating of this indicator is harmonized EU SILC statistical survey. We focus on distribution of poverty and social exclusion in the regions of Slovakia and Czech Republic. We describe current trends for aggregate indicator in Slovakia and Czech Republic and compare our values and trends with others EU countries. The paper uses selected statistical and mathematical models and procedures.

Keywords: Europe 2020 Strategy, At-Risk-of-Poverty, Material Deprivation, Low Work Intensity, Region, EU SILC database.

Estimating Heterogeneous Time-Varying Parameters in Brand Choice Models

Winfried J. Steiner[1], Bernhard Baumgartner[2], Daniel Guhl[3], Thomas Kneib[4]
[1]*Clausthal University of Technology, Dept of Marketing, Germany,*
[2]*University of Osnabrück, Dept of Marketing, Germany,*
[3]*Humboldt University Berlin, Institute of Marketing, Germany,*
[4]*Georg-August-Universität Göttingen, Dept of Statistics and Econometrics, Germany*

Nowadays, modeling brand choice using the multinomial logit model is standard in quantitative marketing. In most applications estimated coefficients representing brand preferences and/or the sensitivity of consumers regarding price and promotional activities are assumed to be constant over time. Marketing theories as well as the experience of marketing practitioners however suggest the existence of trends and/or short-term variations in particular parameters. Hence, having constant parameters over time is a highly restrictive assumption.
We therefore develop a flexible heterogeneous multinomial logit model to estimate time-varying parameters. Both time-varying brand intercepts and time-varying effects of covariates are modeled based on penalized splines, a flexible yet parsimonious nonparametric smoothing technique. The estimation procedure is fully data-driven, determining the flexible function estimates and the corresponding degree of smoothness in a unified approach. In addition, our approach allows for heterogeneity in all parameters.
To assess the performance of the new model, we compare it to models without time-varying parameters and/or without heterogeneity as well as to further benchmark models. Our findings suggest that models allowing for time-varying effects can significantly outperform models assuming constant parameters both in terms of fit and predictive validity.

Keywords: Brand Choice Modeling, Time-Varying Parameters, Heterogeneity, Semiparametric Regression, P(enalized) Splines.

The Feed Consumption and the Piglets' Growth during the Rearing Period Observed in 2015 VS Expected in 2005

Marijan Sviben
Freelance Consultant, Croatia

In the Dr. Manfred Weber's research report published as the professional information by Landesanstalt für Landeswirtschaft und Gartenbau in Bernburg, Sachsen-Anhalt, B.R. of Germany, in November 2015 it has been written that hybrid piglets during the rearing period (from the age of 28 days till the age of 49 days) consumed on an average 393 g of the feed mixture (18 days the starter mixture containing 14.2 MJ of the energy and 17.4% of the crude proteins per kilogramme, 3 days the grower mixture having 14.2 MJ of the energy and 18.4% of the crude proteins per kilogramme) and they gained 282 g a day. The weaners averaged 8.3 kg at the age of 28 days and achieved 9.0 kg at the age of 35 days, 11.0 kg at the age of 42 days and 14.3 kg at the age of 49 days. With those mean values the minimum of the feed mixture required for the maintenance (KG_{FM}) of the piglets' live weight (KG_{LW}) was calculated using the equation $KG_{FM} = KG_{LW}^{5/9} \times 0.0713598$. It was found out that 248 g of consumed feed mixture had been required for the maintenance of the piglets' live weight a day. From 393 g of consumed feed mixture 145 g per day could be used for the production of the piglets' gain. At the International Conference "Modern Genetics, the Nutrition and the Management in the Pig Industry of Serbia" in September 2005 it was shown that 215 g/day of the feed mixture were required for the production of 372 g/day of the gain of piglets, contemporarily bred and held under favourable circumstances, during the period from the age of 28 days till the age of 49 days. In 2015 it was observed that piglets could use 0.6744 of the quantity of the feed mixture required for the production of the gain of live weight. It was possible to conclude that those piglets were able to gain 0.6744 of 372 g, i.e. 251 g a day. The piglets' gain was lessened because the part of consumed feed mixture available for the production of live weight was diminished but observed gain of 282 g/day did 1.1235 of possible magnitude, doing 0.7581 of the gain expected from 2005. That deviation is explainable by the fact that piglets in 2015 consumed approximately 74% of the rate of crude proteins estimated to be suitable for contemporarily bred – hybrid piglets during the rearing period.

Keywords: Piglet, Rearing, Feed Consumption, Growth.

Merging and Diffusion Approximation of Stochastic Epidemic Models

Mariya Svishchuk
Mount Royal University, Canada

A particular challenge in epidemic modelling is the appropriate way to allow for spatial population structure, whereby the rate of contacts between hosts depends on their spatial separation or relative position in a social network. To capture the real complexity of such dynamics, we propose a model of the evolution of epidemic and awareness spreading processes on a different population partitions, mixing patterns and mobility structures. Based on random evolutions, we study merging and diffusion approximation of structured epidemic models.

Keywords: epidemics, merging, diffusion approximation.

Stochastic Modelling and Pricing of Energy Markets' Contracts with Local Stochastic Delayed and Jumped Volatilities

Anatoliy Swishchuk
University of Calgary, Canada

In this talk we study stochastic modelling and pricing of electricity, gas and temperature markets' contracts with delay and jumps, modelled by independent increments processes. The spot price models are based on a sum of non-Gaussian Ornstein-Uhlenbeck processes (including geometric and arithmetic models), describing the long- and short-term fluctuations of the spot dynamics. The models include jumps not only in the spot dynamics but also in the stochastic volatility to describe typical features like spikes of energy spot prices and jumps in the volatility. The basic products in these markets are spot, futures and forward contracts, swaps and options written on these, which will be investigated in our talk.

Keywords: energy markets; stochastic, jump and delayed volatilities; pricing energy contracts; spot, forwards and futures.

Fuzzyfying Labour Market States

Maria Symeonaki
Dept of Social Policy, School of Social and Political Sciences,
Panteion University of Social and Political Sciences, Greece

In this paper the theory of fuzzy logic and fuzzy reasoning is combined with the theory of Markov systems and the concept of a non homogeneous Markov system with fuzzy states is proposed for the measurement of labour mobility. This is an effort to deal with the uncertainty introduced in the estimation of the transition probabilities, especially when social mobility is being measured, and the fact of fuzzy states, in the sense that the categories cannot be precisely measured and are therefore fuzzy. A description of the methodology is outlined and the basic parameters of the system are estimated. Moreover, the proposed methodology is illustrated through the example of measuring transitions between labour market states based on raw data drawn from the EU-LFS survey.

Keywords: Markov systems, Fuzzy Markov systems, Probability of fuzzy events, Fuzzy states, Social mobility**,** EU-LFS.

Acknowledgements: *This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649395 (NEGOTIATE – Negotiating early job-insecurity and labour market exclusion in Europe, Horizon 2020, Societal Challenge 6, H2020-YOUNG-SOCIETY-2014, Research and Innovation Action (RIA), Duration: 01 March 2015 – 28 February 2018).*

A Neuro-Fuzzy Approach to Measuring Attitudes

Maria Symeonaki, Aggeliki Kazani
Panteion University of Social and Political Sciences,
Dept of Social Policy, Greece

The present paper deals with the application of neuro-fuzzy techniques to the measurements of attitudes. The methodology used is analyses and it is illustrated and evaluated on data drawn from a large-scale survey conducted by the National Centre of Social Research of Greece, in order to investigate opinions, attitudes and stereotypes towards the "other" foreigner. The illustration provides a meaningful way of classifying respondents into xenophobic levels, taking also into account other important socio-demographic characteristics, such as age and gender. The methodology moreover provides a way of determining answers provided by respondents that are questionable.

Keywords: Likert scales, attitude measurement, neuro-fuzzy systems.

Labour Market flows in Europe: Evidence from the EU-LFS

Maria Symeonaki, Glykeria Stamatopoulou, Maria Karamessini
Dept of Social Policy, School of Social and Political Sciences,
Panteion University of Social and Political Sciences, Greece

It has been revealed that transition patterns differ considerably across Europe, showing a continuous decrease in the probabilities of remaining into full-time employment and an increase in the unemployment rates or unstable employment. This trend is more common for young people, as they are in a more disadvantaged position, facing much more turnover between employment, unemployment and inactivity compared to the older workers. The present paper presents an analysis on the labour market dynamics in Europe. Raw data drawn from the European Union Labour Force Survey (EU-LFS) are used to show similarities and differences in the labour market flows across European countries for the years 2014 and 2015. Methodologically, we use the theory of non-homogeneous Markov system. More particularly, the population is stratified according to the labour market status and transition probability matrices are generated by educational level and gender, where possible to show movements between the labour market states of employment, unemployment and inactivity and vice versa. Indices are calculated for two age cohorts, in order to compare labour market patterns between young and older people.

Keywords: Labour market flows, Markov systems, transition probabilities, EU-LFS, Europe.

Acknowledgements: *This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649395 (NEGOTIATE – Negotiating early job-insecurity and labour market exclusion in Europe, Horizon 2020, Societal Challenge 6, H2020-YOUNG-SOCIETY-2014, Research and Innovation Action (RIA), Duration: 01 March 2015 – 28 February 2018).*

On the Measurement of Early Job Insecurity in Europe

Maria Symeonaki, Glykeria Stamatopoulou, Maria Karamessini
Dept of Social Policy, School of Social and Political Sciences,
Panteion University of Social and Political Sciences, Greece

In the present paper the estimation of different indicators that can be used in order to capture the extent and forms of early job insecurity is studied. This specific matter has been receiving increasing research and policy attention throughout the two last decades. The present study proposes a new index for measuring the degree of early job insecurity on

the basis of the estimation of the transition probabilities between labour market states and school-to-work transitions, with raw data drawn from the latest available EU's Labour Force Survey (EU- LFS), i.e. for the year 2014 and 2015. This indicator reflects a more meaningful way of estimating labour mobility or fluidity for young individuals aged between 15 and 29 adjusting the well-established mobility indices to the labour market area. Thus, an attempt is made basically to provide a new measure for early job insecurity, connecting it also to school-to-work transition probabilities, that captures the extent of early job insecurity.

Keywords: Labour market flows, Markov systems, transition probabilities, EU-LFS, Europe.

Acknowledgements: *This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649395 (NEGOTIATE – Negotiating early job-insecurity and labour market exclusion in Europe, Horizon 2020, Societal Challenge 6, H2020-YOUNG-SOCIETY-2014, Research and Innovation Action (RIA), Duration: 01 March 2015 – 28 February 2018).*

FISS- The Factor Based Index of Systemic Stress

Tibor Szendrei, Katalin Varga
Central Bank of Hungary, Hungary

Tracking and monitoring stress within the financial system is a key component for financial stabil-ity and macroprudential policy purposes. Financial stress measures are important as forward looking indicators to signal potential vulnerabilities in the market, enabling policy makers to take corrective measures in time, minimizing the impact on the real economy. This paper introduces a new measure of the contemporaneous stress within the Hungarian financial system named Fac-tor based Index of Systemic Stress (FISS). The aim was to capture the common components of financial stress that optimally compress the information available in high-dimensional data. Its statistical design is a dynamic Bayesian factor method. The main methodological innovation of the FISS is the ability to fully capture information contained in persistent high-frequency data, namely the usage of common stochastic trends as factors. We determine the optimal linear com-bination of factors resulting in the final index with Information Value methodology. Applied to Hungarian data the FISS is planned to be a key element of the macroprudential toolkit.

Keywords: financial system, financial stability, systemic risk, financial stress index, macro-financial linkages.

Random Effects Models in Item Analysis: Exploring Applications in Autobiographical Memory

Angela Tagini
Dept of Psychology, Universita Milano-Bicocca, Italy

Random effects models have found application in several areas of behavioral studies, when data are clustered. In particular, in the case of more observations for each person, repeated measures are clustered within subjects. Two situations occur. In the first one, the same measurement is repeated in time, leading to grow curves modeling. In the second one, a certain set of stimuli and/or items are submitted top to respondents, therefore the model applies to the set of items/stimuli within the same person. The heterogeneity within and between respondents are typical sources of variation in response data that needs to be accounted for in an appropriate statistical response model. The typical approach to study items in a scale or a test is Item Response Theory (IRT): the analysis produces a score and a standard error for each case. On the other hand, item responses data can be recognized as a clustered or nested structure. Namely, an item response model is a non linear mixed effects model. Observations from each respondent are mutually dependent, as they share a common underlying trait, for instance ability. First level (L1) units are responses to items, subjects represent level 2 (L2) units. In the case of the two parameter model and of a dichotomous response:

$$h\left(E\left(y_{ik}\backslash\beta_i - \delta_j\right)\right) = \alpha_j\beta_i - \delta_j \qquad (1)$$

where $h(.)$ is the logit link function, i indexes subjects, $i = 1, 2, .., n, j,$ indexes items and $j = 1, 2, .., K.$ Both item and person parameters are random. Expression (1) is a non linear mixed effects model and the correlations of the L1 item responses are described by the random person parameters. Mixed random effects models are flexible, since they provide, in addition to item and person parameters, test for differential item functioning (DIF). The applicative advantages of this methodological frame are explored with respect to a specific behavioral topic, autobiographical memory, within an ongoing research were data are being collected in line with a coherent and sound behavioural theoretical approach. The application extends the model in (1), with a dichotomous response, to the case of a politomous response, in the so-called Partial Credit (PC) model.

Keywords: item, random effect, nested, latent.

Experiment with a Survey-Based Election to the Student Parliament of the Karlsruhe Institute of Technology

Andranik Tangian
Institute of Economic and Social Research (WSI)
at the Hans Böckler Foundation, Germany

Since voters are often swayed more by the personal image of politicians than by party manifestos, they may cast votes that are in opposition to their policy preferences. This results in the election of representatives who do not correspond exactly to the voters' own views. An alternative voting procedure to avoid this type of election failure is proposed in some previous publications by the author. It is based on the approach implemented in internet voting advice applications, like the German Wahl-O-Mat, which asks the user a number of questions on topical policy issues; the computer program, drawing on all the parties' answers, finds for the user the best-matching party, the second-best-matching party, etc. Under the proposed alternative election method, the voters cast no direct votes. Rather, they are asked about their preferences on the policy issues as declared in the party manifestos (Introduce nationwide minimum wage? Yes/No; Introduce a speed limit on the motorways? Yes/No, etc.), which reveals the balance of public opinion on each issue. These embedded referenda measure the degree to which the parties' policies match the preferences of the electorate. The parliament seats are then distributed among the parties in proportion to their indices of popularity (the average percentage of the population represented on all the issues) and universality (frequency in representing a majority). This paper reports on an experimental application of this method during the election of the Karlsruhe Institute of Technology Student Parliament on July 4–8, 2016. The experiment shows that the alternative election method can increase the representativeness of the Student Parliament. We also discuss some traits and bottlenecks of the method that should be taken into account when preparing elections.

Keywords: Election, voting, parliament, policy representation, representative democracy.

On the Kernel Hazard Rate Estimation under Association and Truncation

Abdelkader Tatachak, Zohra Guessoum
Laboratory MSTD, Dept of Probability and Statistics, USTHB, Algeria

Survival analysis is the part of statistics, in which the variable of interest (lifetime) may often be interpreted as the time elapsed between two

events and then, one may not be able to observe completely the variable under study. Such variables typically appear in a medical or an engineering life test studies. Among the different forms in which incomplete data appear, random left truncation is a common one.
The main results stated in the present work deal with the asymptotic analysis of a kernel estimator for the hazard rate function under left truncated and associated model. For this, we first establish strong uniform consistency rates and then, we provide simulation results to evaluate the finite-sample performances of the proposed estimator.

Keywords: Associated data, Hazard rate function, Lynden-Bell estimator, Random left truncation, Strong uniform consistency rate.

The Infinite Horizon Ruin Probability in the Cramer-Lundberg Model with Two-Sided Jumps

Eleftherios Theodosiadis, George Tsaklidis
Dept of Mathematics, Aristotle University of Thessaloniki, Greece

Given the Cramer-Lundberg model with double sided jumps that are expressed by two stochastically independent compound Poisson processes, we study the behavior of the related ruin probability function (see also Vidmar (2016), and Pertsinidou (2017)). We derive an integral equation concerning the ruin function using analytic methods, only. Furthermore, we provide a procedure to find the ruin function when the jumps of the model follow some special, continuous distributions, under certain conditions and we ask whether these conditions can be practically relaxed.

Estimation of Two Sided Asset Return Jumps via Constraint Kalman Filtering

Ourania Theodosiadou, George Tsaklidis
Dept of Mathematics, Aristotle University of Thessaloniki, Greece

The positive and negative jumps underlying the daily log returns of the Nasdaq index are estimated via constraint Kalman filtering. These jumps are determined by the arrival of positive and negative news in the market and are hidden (i.e., not observable). In order for the jumps to be non-negative, their probability density functions are appropriately truncated according to the non-negativity constraints and then the associated prior and posterior state estimates are derived. Finally, the fitting of this model to the empirical data is examined.

Frost Prediction in Apple Orchards Based upon Time Series Models

Monika A. Tomkowicz[1], Armin O. Schmitt[2]
[1]*Faculty of Science and Technology, Free University of Bozen/Bolzano, Italy,*
[2]*Breeding Informatics, Faculty of Agricultural Sciences, Georgia-Augusta-University of Göttingen, Germany*

The scope of this work was to calculate a frost forecast model for South Tyrol in Italy using weather data of the past 20 years which were recorded by 150 weather stations located in this region. Accurate frost forecasting should provide growers with the opportunity to prepare for frost events in order to avoid frost damage. The radiation frost in South Tyrol occurs during the so-called frost period, i.e. in the months of March, April and May during calm nights between sunset and sunrise. In case of a frost event, the farmers should immediately switch on water sprinklers. The ice cover which is built on the trees protects the buds or blossoms from damage. Based on the analysis of time series data, the proposed linear regression and ARIMA models could be compared and evaluated. The best result provided the ARIMA model, achieving in case of forecast of 95% confidence intervals the desired value of 1.0 for the recall. This means that all frost cases could be correctly predicted. Despite the encouraging results, the rate of the false positives is high, which needs further investigations (e.g., testing VARIMA models, which are a multivariate extension of ARIMA models). The graphical illustration of the 95% confidence intervals of the ARIMA model forecast should be very helpful in frost prediction and could be integrated in the electronic monitoring system that permits intelligent forecasting of frost weather phenomena.

Keywords: Frost Forecast, ARIMA Models, Time Series Prediction Models.

Familys' Transfers to the Divorced Elderly

Ekaterina Tretyakova
The Russian Presidential Academy of National Economy and Public Administration (RANEPA), Russia

Within the aging of population the question of care and financial support of elderly become increasingly relevant. Meanwhile the spread of divorce and separation becomes a very strong trend in the modern society. Divorced elderly cannot get support from their partner, they do not work anymore and thus their income consists only of pensions which are often not enough to provide the adequate standard of living. Transfers were significantly described in scientific literature (P. Samuelson, R. Baker),

but not from point of view of marital status. In this research, based on data from the «Comprehensive monitoring of living conditions of the population» conducted in Russia in 2014, author examines how children in Russia help their old parents and how this help varies depending on gender and marital status.

The main factor here is the fact that life expectancy of men about 11 years less than of women. Due to this fact the rate of lonely women above 55 years old is much higher than the percentage of lonely men. Also men tend to remarriage easier than women, that's why they more often find a new family. It is a common practice in Russia that after divorce or separation children stay with mother and this make the remarriage for women more difficult. On the other hand this tight connection with children provides women more support than men.

For example through all marital statuses the biggest gender differences in material transfers are observed among divorced and separated people: the percent of men who get material support from their children is much lower than the percent of women (19,5% against 35,9% among divorced and 10% against 37,9% among separated). On the other hand the average income of men during working period of life is higher than income of women, due to this fact they more often have savings.

Another question is intangible help – housekeeping and care during illness – which can not be earned or saved and depends only on the relationship with children. From this point of view divorced and separated elderly men is the most vulnerable group: only one third of them get help from children.

At last when we analyze elderly who do not get help from their children because of the termination of the relationship, we see that "separated group" is presented only by men and 80% the "divorced group" is presented by men only.

To sum up the results Russia there is a big gap between elderly man and women in getting help from their children. Tight connection between children and mother after divorce or separation leads to more intensive transfers from this children when the woman gets retirement. Men make more savings than women which can help them to overcome the lack of the material support from their children, but they can not help to overcome the lack of intangible help. This problem can be solved with the development of institutions with which men can get such services, but unfortunately this sphere is still not high-developed in Russia.

Keywords: transfers, houshold, family, divorce.

A Local Approach based on Risk Measures for the Hedging of Variable Annuities

Denis-Alexandre Trottier, Frédéric Godin, Emmanuel Hamel, Richard Luger
Concordia University, Canada

We present a simple local hedging approach based on risk measures for the hedging of variable annuities in the presence of equity risk, interest rate risk and basis risk. The hedging strategy is obtained by minimizing risk with respect to next-period's cash flow injection within the hedging portfolio by the insurer. Taylor expansion based approximations are used to improve the tractability of the approach by reducing the problem's dimensionality. The impact of basis risk on capital requirements in quantified. The hedging performance of our approach is compared to industry benchmarks such as Fund Mapping Regressions.

Keywords: Hedging, Variable Annuities, Risk Measures, Basis Risk.

Nano-Sols Shelf-Life Prediction via Accelerated Degradation Model

Sheng-Tsaing Tseng
National Tsing-Hua University, Taiwan

In order to provide timely product's lifetime information to the customers, conventionally, manufacturers usually use temperature (or voltage) as an accelerating variable for shortening life testing time. Based on well-known life-stress relationship (such as Arrhenius reaction or inverse power model), it allows us to extrapolate the lifetime of highly-reliable products at a normal used condition. In this talk, however, we will present a real case study that the shelf-life prediction of nano-sol products can be successfully obtained by adopting pH value as an accelerating variable. An accelerated profile-degradation model is proposed to describe the time-evolution of the particle size distributions under three different pH values. Then, we can analytically construct the confidence interval for the shelf-life of nano-sol products under its normal use condition.

Keywords: nano-sol, accelerated profile-degradation model, shelf-life prediction.

Adaptive MCMC for Multivariate Stable Distributions

Ingrida Vaičiulytė
Business and Technologies Faculty, Šiauliai State College, Lithuania

Markov chain Monte Carlo adaptive methods by creating computationally effective algorithms for decision-making of data analysis with the given accuracy are analyzed in this paper. The task for estimation of parameters of the multivariate stable symmetric vector law which is constructed in hierarchical way is described and solved in this paper. To create the adaptive MCMC procedure, the sequential generating method is applied for Monte Carlo samples, introducing rules for statistical termination and for sample size regulation of Markov chains. Statistical task, solved by this method, reveal characteristics of relevant computational problems including MCMC method.
Effectiveness of the MCMC algorithms is analyzed by statistical modeling method, constructed in the paper. Tests made with financial data of enterprises, confirmed that numerical properties of the method correspond to the theoretical model. Tests of algorithms have shown that adaptive MCMC algorithm allows obtaining estimators of examined distribution parameters in lower number of chains, and reducing the volume of calculations approximately two times. The algorithms created in this paper can be used to test the systems of stochastic type and to solve other statistical tasks by MCMC method.

Keywords: Monte Carlo method, EM algorithm, maximum likelihood method, stable distributions, stochastic optimization.

Bibliometric Variables Determining the Quality of a Dentistry Journal

Pilar Valderrama, Manuel Escabias,
Evaristo Jiménez-Contreras, Alberto Rodríguez-Archilla
Dept of Statistics, University of Granada, Spain

On the basis of a review of the main journals in the field of Dentistry Science indexed in Journal of Citation Report (JCR) a set of variables, both quantitave such as number of papers per issue, frequency, H-index of authors or number of keywords, and categorical ones such as open access, electronic submission, suitability to CONSORT guide (only for clinical trials) or tables and figures presentation, with potential influence in the quality of those journals, measured by means of the impact factor and the eigenfactor, has been selected. Afterwards, by sampling more than 40 journals in this field, the most significant ones have been chosen

by means of an ordinal regression model. All the calculations have been implemented by R programm.

Keywords: Impact factor, Eigenfactor, Ordinal regression, Dentistry journal.

Are the Leverage Point the most Terrible Problem in Regression?

Jan Ámos Víšek
Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic

When studying property of recently proposed S-weighted estimators (Víšek (2015), (2016)) it appeared that the most previously studied (highly) robust estimators can (and typically) have problems with the outliers when in data are also good leverage points. Of course, it depends on the topology of data – the outliers can't be very far from the bulk of data. The S-weighted estimators are generalization of the Least Weighted Squares estimators as well as of S-estimators. They look for the minimal value of estimator of the standard deviation of disturbances but they also employ weights to depress the influential points which can have improper affect on the estimation of underlying regression model. Due to the fact that the weights are assigned to the observations by the method itself, S-weighted estimators can avoid the problems which the other methods (can) suffer. This malfunction of some estimators is a consequence of the fact that they look (implicitly or explicitly) for the minimal value of estimator of the standard deviation of disturbances in a way which is too much focused on this minimization, not allowing to recognize the information that the good leverage points bring. S-weighted estimators - contrary to S-estimators, W-estimators or M-estimators were able properly find the outliers and utilize also the information offered by the good leverage points. Due to the fact that S-weighted estimators were able to employ efficiently the information contained in the leverage points, their empirical mean square errors (in simulations) were (much) smaller than the mean square errors of other estimators.

Keywords: Robust estimation of regression model, S-estimators, the Least Weighted Squares, S-weighted estimators.

References:

[1]. Víšek, J. Á. (2015): S-weighted estimators. *Proc. of the 16th Conference on the Applied Stochastic Models, Data Analysis and Demographics 2015, ed. Christos H. Skiadas, 1031–1042.*

[2]. Víšek, J. Á. (2016): Representation of SW-estimators. *Proc. 4th Stochastic Modeling Techniques and Data Analysis International Conference with Demographics Workshop, SMTDA 2016, ed. Christos H. Skiadas, 425–438.*

A Generalized Distribution Family of the Freund Bivariate Exponential Model

Juana-Maria Vivo[1], Manuel Franco[1], Debasis Kundu[2]
[1]*Dept of Statistics and Operations Research, University of Murcia, Spain,*
[2]*Dept of Mathematics and Statistics, Indian Institute of Technology, India*

Recently, an extension of the Freund's bivariate exponential (FBE) load share distribution has been provided by Asha, Krishnan and Kundu (2016) and called the extended Freund's bivariate (EFB) distribution. The Freund (1961) model is based on the lifetime distributions of the two-component parallel redundant systems whose components have exponential distributions, and it was extended through Weibull components by Lu (1989). In the recent EFB model, Asha et al. (2016) have considered components having proportional failure rate models with an underlying distribution.
In this work, a generalization of the EFB (GFB) distribution is derived taking into account how failure rates change after failing a component. Specifically, when a component fails, the baseline distribution function of the remaining working component could also change, since the instantaneous failure probability (failure rate) is expected to be modified due to overload of the surviving component.

Keywords: Freund exponential model, bivariate distribution, failure rate, parallel system.

References:

[1]. Asha, G., Krishnan, J.K.M. and Kundu, D. (2016). An extension of the Freund's bivariate distribution to model load sharing systems. American Journal of Mathematical and Management Sciences.
[2]. Freund, J.E. (1961). A bivariate extension of the exponential distribution. Journal of the American Statistical Association, 56, 971-977.
[3]. Lu, J.C. (1989). Weibull extensions of the Freund and Marshall-Olkin bivariate exponential models. IEEE Transactions on Reliability, 38, 615–620.

Grouping Property and Relative Importance of predictors in Linear Regression

Henri Wallard
Ipsos Science Center, France

The quantification of the importance of predictors on a response variable has been a subject of research in several fields such as biostatistics, psychology, economics or market research. Regression analysis may be used for that purpose but estimating importance of predictors via (standardized) regression coefficients is often not adequate because of

the presence of correlations between these variables. Therefore, alternative methods have been considered.
Grouping property is said to be respected when estimators of importance tend to equate for highly correlated predictors. While standardized regression coefficients for instance do not respect grouping property, we will analyze the respect of grouping property for some methods used to quantify relative importance through decomposition of the explained variance such as Fabbris, Shapley Value, Genizi-Johnson and CAR scores (Correlation-Adjusted Correlation). These analyses will be based on theoretical demonstration and illustrated with datasets.
CAR scores have been recommended as estimators of importance of predictors in the field of biostatistics and been justified by the respect of grouping property. Quite contrary we will show theoretically and on examples that CAR scores do not respect this property. We will explain in return why some other variance decomposition methods do respect grouping property. We will also discuss quantification of importance using random forests in the perspective of grouping property.
Lastly, we will formulate recommendations for estimation of the relative importance of predictors.

Keywords: Variance decomposition, multiple regression, CAR scores, random forests.

On GARCH Models with Temporary Structural Changes

Norio Watanabe, Fumiaki Okihara
Chuo University, Japan

When an economic shock like the Lehman Crisis occurred, it is expected to investigate its influence based on economic time series. The intervention analysis by Box and Tiao is a method for such a purpose. Most of the intervention analyses are based on ARIMA models, but some are on GARCH models. The GARCH models have been developed for analyzing time series of stock returns. Usually the expected value function of a GARCH model is assumed to be constant. However, this assumption is not appropriate when a time series includes a varying trend. Our first purpose is to propose a trend model, which can be easily taken in intervention analysis.
Furthermore we generalize this model for an intervention analysis on both of trend and volatility in a GARCH model. An identification method is also provided and is evaluated by simulation studies. Usability of the proposed model is demonstrated by applying to real stock returns.

Keywords: Intervention analysis, stock return, trend.

Operator Models in Theory of Branching Random Walks and their Applications

Elena Yarovaya[1,2]
[1]*Dept of Probability Theory, Lomonosov Moscow State University, Russia,*
[2]*Steklov Mathematical Institute, Russia*

Stochastic processes with generation and transport of particles are used in different areas of nature sciences: statistical physics, chemical kinetics, population dynamics etc. Nowadays it is commonly accepted to describe such processes in terms of branching random walks. Branching random walk is a stochastic process which combines the properties of a branching process and a random walk. Behavior of branching random walks in many ways is determined by properties of a particle motion and a dimension of the space in which the particles evolve. We consider continuous-time branching random walks on multidimensional lattices with a finite set of the particle generation centers, i.e. branching sources. The description of a random walk in terms of Green's function allows us to offer a general approach to investigation of random walks with finite as well as with infinite variance of jump. The main object of study is the evolutionary operator for the mean number of particles both at an arbitrary point and on the entire lattice. The existence of positive eigenvalues in the spectrum of an evolutionary operator results in the exponential growth of the number of particles in branching random walks, called supercritical in such a case. For supercritical branching random walks, it is shown that the amount of positive eigenvalues of the evolutionary operator, counting their multiplicity, does not exceed the amount of branching sources on the lattice with positive intensity, while the maximum of these eigenvalues is always simple. We demonstrate that the appearance of multiple lower eigenvalues in the spectrum of the evolutionary operator can be caused by a kind of `symmetry' in the spatial configuration of branching sources.

Keywords: Operator Models, Spectral Analysis, Branching Random Walks, Limit Theorems, Population Dynamics.

Acknowledgements: *This research is supported by Russian Science Foundation project no. 14-21-00162.*

Investigating Southern Europeans' Perceptions of Their Employment Status

Aggeliki Yfanti[1], Catherine Michalopoulou[1], Aggelos Mimis[2], Stelios Zachariou[3]
[1]Dept of Social Policy, Panteion University of Social and Political Sciences, Greece, [2]Dept of Economic and Regional Development, Panteion University of Social and Political Sciences, Greece, [3]Hellenic Statistical Authority, Greece

The European Union Labour Force Survey (EU-LFS) measurement of the employment status is based on a synthesized economic construct computed according to the ILO conventional definitions of employed, unemployed and inactive. Since the late 2000s, a variable measuring people's perceptions of their employment status has been included in the EU-LFS questionnaire as it is used in all large-scale sample surveys, i.e. one of the occupational background variables. These measurements are not comparable and their results will differ since a composite economic construct would normally deviate from people's perceptions. The purpose of this paper is, by obtaining a social "profile" of agreement and disagreement between Southern Europeans' self-declared perceptions of their employment status and the ILO conventional definitions, to investigate whether or not conflicting and coinciding perceptions differ overtime within-nations and cross-nationally. The analysis is based on the 2008-2014 annual datasets for Greece, Italy, Portugal and Spain. The results are reported for the age group 15-74 so as to allow for comparability with the ILO conventional definition of unemployment.

Keywords: Employment status, ILO, EU-LFS, Southern Europe.

Estimation the Key Value of Shift Cipher by Neural Networks – A Case Study

Eylem Yucel, Ruya Samli
Dept of Computer Engineering, Istanbul University, Turkey

Security is an extensive concern in information and data systems. One of the most important ways for keeping information secure is cryptology which provides various algorithms to perform substitutions and transformation on the original text to produce unintelligible ciphertext. There are many algorithms in cryptology field. The public and private key pair comprise of two uniquely related cryptographic keys. In private key cryptography, the key is secret and it uses the same cryptographic keys for both encryption and decryption processes. In public key cryptography, encryption and decryption parts use different keys and one

of them is private while the other is public. Any person can encrypt a message using the public key of the receiver, but such a message can be decrypted only with the receiver's private key. When a cipher method is applied to a plaintext, first of all the letters must be transformed to numbers as:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Shift Cipher is one of the methods in private key cryptography and involves the shifted letters of all the letters in plaintext by a specific step number which is the key of the algorithm and constitutes ciphertext. Some samples of plaintext and ciphertext in terms of English alphabet can be seen in the table below.

Key = 0

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Key =3 (Caesar Cipher)

D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

Key = 10

K L M N O P Q R S T U V W X Y Z A B C D E F G H I J

Neural networks is a modelling and estimating method which is used in many fields. In this study, we use neural networks to find the private key and cryptanalysis of Shift Cipher methodology in a specific case study. The results show that neural networks is an appropriate method for Shift Cipher cryptanalysis and in future work we will observe if the cryptanalysis of other cryptosystems can be done by neural networks or not.

Keywords: Neural Networks, Private Key Cryptography, Shift Cipher.

Dealing with Inaccuracies and Limitations of Empirical Mortality Data in Small Populations

K. Zafeiris[1], A. Kostaki[2]

[1]Laboratory of Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece, [2]Dept of Statistics, Athens University of Economics and Business, Greece

This paper provides a description of the most typical problems and limitations affecting mortality data of small populations, discusses their consequences in estimating age-specific mortality patterns, and also proposes methodology for overcoming them. In this context a theoretically consistent though computationally simple technique for minimizing random variations in age-specific death counts is proposed and demonstrated.

Health Estimates for some Countries of the Rapid Developing World

K. N. Zafeiris[1], C. H. Skiadas[2]
[1]Laboratory of Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece, [2]ManLab, Dept of Production Engineering and Management, Technical University of Crete, Greece

The recent developments of health were studied in the emerging economies of Brazil, Russia, India, Indonesia and China. Data come from the World Health Organization in the form of abridged life tables. These tables were unabridged into full ones with the UNABR application of the MORTPAK (4.3) software which has been created by the United Nations Population Division for the needs of mortality analysis. The extended life tables were used subsequently in order for the First Exit Time theory to be applied and a healthy life expectancy to be estimated. Firstly, results indicate a general trend of improvement in these countries. Secondly, they are in accordance with the findings of WHO and its related teams concerning healthy life expectancy, which has been estimated on the basis of a totally different methodology; thus First Exit Time theory application constitutes a rather parsimonious and efficient technique for the estimation of relevant measures.

Limit Theorems for Compound Renewal Processes: Theory and Applications

Nadiia Zinchenko
Dept of Informatics and Applied Mathematics, Nizhyn State Mukola Gogol University, Ukraine

We consider a few classes of strong limit theorems for compound renewal processes (random sums, randomly stopped sums) D(t) = $S(N(t)) = \sum_{i=1}^{N(t)} x_i$ under various assumptions on the renewal counting process N(t) and random variables $\{x_i, i \geq 1\}$, which summarize authors previous results obtained during last five years or so. First of all we present sufficient conditions for strong (a.s.) approximation of D(t) by a Wiener or α-stable Lévy process under various dependent and moment conditions on summands, mainly focused on the cases of independent, weakly dependent, φ-mixing and associated r.v. On the next step a number of integral tests for investigation the rate of growth of the process D(t) and it's increments D(t+a(t)) - D(t), when a(t) grows as $t \to \infty$, are proposed. As a consequence various modifications of the LIL and Erdös-Rényi-Csörgő-Révész-type SLLN are obtained. Useful

applications in financial mathematics, queuing and risk theories are investigated; particularly, non-random bounds for the rate of growth and fluctuations of the risk processes in classical Cramer-Lundberg and renewal Sparre Andersen risk models are discussed as well as the case of risk models with stochastic premiums.

Keywords: Compound Renewal Process, Random Sum, Limit Theorem, Strong Approximation, Integral Tests, Queuing Theory, Risk Process.

The Score Correlation Coefficient

Zdeněk Fabián
Institute of Computer Science, Czech Republic

The usual measure of linear dependence of random variables X and Y is the correlation coefficient. Its currently used estimates are independent of marginal distributions of both X and Y. Usually, it is supposed that:
i) if X and Y have light-tailed distributions, the best estimate is the Pearson correlation coefficient
ii) if some of them has contaminated light-tailed distribution, the best estimate is some version of robust correlation coefficients
iii) if some of them has heavy-tailed distribution, the best estimate is a rank correlation coefficient.
Recently, Fabián (2001) generalized the concept of the scalar score fuction of distributions having support the entire R and location parameter for any continuous distribution F(x; ɵ) with arbitrary support and arbitrary vector ɵ of parameters. This function, T(x; ɵ), say, is considered to be the likelihood (Fisher) score with respect to the (newly introduced) 'center' of distribution F. T(x; ɵ) of some currently used models is identical with the actual likelihood with respect to a ('central') component of ɵ.
Using this 'relative influence of observations' function makes possible to study distribution-dependent score correlation coefficients and to verify assumptions i)-iii).
These assumptions were partly confirmed by our simulation experiments. The score correlation coefficient seems to be the best estimate of linear dependence of X and Y only if they have skewed, 'medium' heavy-tailed distributions with Pareto-type tail and a reasonably small value of the generalized coefficient of variation. Moreover, we found that if the data stem from heavy-tailed distributions, all the estimates of the correlation coefficient depend to the certain extent on the estimated value.

Keywords: score function, heavy tails.

Reference: Fabián Z. (2001). Induced cores and their use in robust parametric estimation. Comm. Statist. Theory -Methods, 30, 537-556.

Authors Index

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

V

W

X

Y

Z