Book of Abstracts

4th Stochastic Modeling Techniques and Data Analysis International Conference with 5th Demographics Workshop

SMTDA2016

Editor

Christos H Skiadas

June 1 - 4, 2016

University of Malta, Valletta, Malta

Imprint
Book of Abstracts of the 4th Stochastic Modeling Techniques and Data Analysis International Conference with the 5th Demographics Workshop
Valletta, Malta: 1-4 June, 2016
Published by: ISAST: International Society for the Advancement of Science and Technology.
Editor: Christos H Skiadas

Book ISBN: 978-618-5180-14-0
e-Book ISBN: 978-618-5180-15-7

Preface

It is our pleasure to welcome the guests, participants and contributors to the International Conference (SMTDA 2016) on Stochastic Modeling Techniques and Data Analysis and (DEMOGRAPHICS2016) 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.

SMTDA Conference and DEMOGRAPHIC 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.

SMTDA 2016 and DEMOGRAPHICS 2016 focus in expanding the development of the theories, the methods and the empirical data and computer techniques, and the best theoretical achievements of the Stochastic Modeling Techniques 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 Book of Abstracts. Special thanks to the Plenary and Keynote Speakers, the Session Organisers, the Scientific Committee, the ISAST Committee and Yiannis Dimoticalis, the Conference Secretary Mary Karadima and all the members of the Secretariat. Many thanks to Lino Sant and Mark Anthony Caruana for their initiative and support to organise this event in the University of Malta and the officials of the University. Lucienne M Bageja, Operations and events Manager, was very important in scheduling and organising the various aspects of a complex system as is the SMTDA Conference. Veronica Barbara helped in many details along with the other people from the Conference Unit.

May 2016

Raimondo Manca, Rome, Italy
Sally McClean, Coleraine, UK
Christos H. Skiadas, Chania, Crete, Greece
Mariano Valderrama, Granada, Spain

SCIENTIFIC COMMITTEE

SMTDA Committee

Local Organizers

Plenary Talks

Jaromir Antoch
Charles University in Prague, Czech Republic
Estimation of Fiber System Orientation Local Approach Based on Image Analysis

Narayanaswamy Balakrishnan
Distinguished University Professor of Mathematics & Statistics, McMaster University
Hamilton, Ontario, Canada
Flexible Cure Rate Models and Applications

Claude Lefèvre
Professeur Département de Mathématique, Probabilités et Modélisation Stochastique
Université Libre de Bruxelles, Belgium
Appell polynomials, first crossing time and insurance risk models

Hans - J. Lenz
Inst. f. Statistik und Ökonometrie, Inst. f. Wirtschaftsinformatik
Freie Universität Berlin, Germany
Knowledge based Tax Fraud Investigation

Jean-Marie Robine
Research Director, French National Institute of Health and Medical Research
Chair, Department of Biostatistics, University of Montpellier 1, France
Can We Use the Highest Reported Age at Death as Proxies of the Maximum Life Span?

Contents **Page**

BOOK OF ABSTRACTS

Stochastic Modeling Techniques and Data Analysis International Conference with Demographics Workshop

Plenary and Keynote Talks

Estimation of Fiber System Orientation Local Approach Based on Image Analysis

Jaromir Antoch

Dept of Probability and Mathematical Statistics
Charles University in Prague, Czech Republic

Analysis of materials often includes measurement of structural anisotropy or directional orientation of object systems. To that purpose the real-world objects are replaced by their images, which are analyzed, and the results of this analysis are used for decisions about the product(s). Study of the image data allows to understand the image contents and to perform quantitative and qualitative description of objects of interest. This lecture deals particularly with the problem of estimating the main orientation of fiber systems. First we present a concise survey of the methods suitable for estimating orientation of fiber systems stemming from the image analysis. The methods we consider are based on the two-dimensional discrete Fourier transform combined with the method of moments. Secondly, we suggest abandoning the currently used global, i.e. all-at-once, analysis of the whole image, which typically leads to just one estimate of the characteristic of interest, and advise replacing it with a \local analysis". This means splitting the image into many small, non-overlapping pieces, and estimating the characteristic of interest for each piece separately and independently of the others. As a result, we obtain many estimates of the characteristic of interest, one for each sub-window of the original image, and - instead of averaging them to get just one value - we suggest analyzing the distribution of the estimates obtained for the respective sub-images.
Proposed approach seems especially appealing when analyzing, e.g., nanofibrous layers and/or nonwoven textiles, which may often exhibit quite a large anisotropy of the characteristic of interest.

Key words and phrases. Fiber system, digital image, Fourier analysis, covariance matrix analysis, moments of image, nanofibers layers, histogram, kernel density estimator.
Lecture is based on joint work with M. Tunak, J. Kula and J. Chvojka.

References

1. Enomae T., et al. Nondestructive determination of fiber orientation distribution of fiber surface by image analysis. Nordic Pulp and Paper Research J. 21, 253-259, 2006.
2. Pourdeyhimi R.B. and Davis D.H. Measuring Fiber Orientation in Nonwovens Part III: Fourier Transform. Textile Research Journal 67, 143-151, 1997.
3. Rataj J. and Saxl I. Analysis of planar anisotropy by means of Steiner compact. J. of Applied Probability 26, 490-502, 1989.
4. Tunak M., Antoch J., Kula J. and Chvojka J. Estimation of fiber system orientation for nonwoven and nanofibrous layers: Local approach based on image analysis. Textile Research Journal. (First-on-line December 12, 2013. DOI: 10.1177/0040517513509852)
5. Tunak M. and Linka A. Planar anisotropy of fiber system by using 2D Fourier transform. Fibers and Textiles in Easter Europe 15, 86-90, 2007.

Flexible Cure Rate Models and Applications

N. Balakrishnan

McMaster University, Canada

In this talk, I will introduce some cure rate models and destructive cure rate models and describe likelihood inferential methods and model discrimination procedures for these models. After presenting some simulation results, I shall illustrate the models and methods with data from melanoma trials.

Appell Type Polynomials, First Crossing Times and Risk Processes

Claude Lefèvre

Département de Mathématique, Université Libre de Bruxelles, Belgium

This work is concerned with two remarkable families of polynomials, related but different, the well-known Appell polynomials and the less-known Abel-Gontcharoff polynomials.
The two families of polynomials are first considered in their univariate version. They are interpreted in terms of the joint right-tail or left-tail distributions of order statistics for a sample of uniforms. This allows us to determine the first-crossing times of a special class of point processes in an increasing upper or lower boundary. They are then defined with randomized parameters that correspond to sums or products of

independent random variables. Thanks to this extension, we can obtain the first crossing times for two important topics in risk theory: the finite-time ruin probability for an insurance model and the final size distribution for an epidemic model.
The polynomials are next examined in their multivariate version, which is little standard in the literature. An interpretation of these polynomials is provided through a random walk on a point lattice with lower or upper bounds on the edges. By randomizing the involved parameters, we are then able to deal with the ruin probability for a multiline insurance model and the final epidemic size for a multigroup model.

Knowledge Based Tax Fraud Investigation

Hans - J. Lenz

Freie Universität Berlin, Germany

Tax Fraud is a criminal activity done by a manager of a firm or by a tax payer who intentionally manipulates tax data to deprive the tax authorities or the government of money for his own benefit. Tax fraud is a kind of Data Fraud, and happens every time and everywhere in daily life of households, business, and science, health care or even in religious communities etc.
In times where the hype about “Big data” is dominating tax fraud detection is intrinsically related to the “Small data” area, cf. “sparse data”. Starting with a prior suspicion the power of human creativity is needed for a step-by-step investigation of the private or firm’s environment hunting for the tax liability materialized in various forms like cash, foreign currency, gold, equities etc. The point is that in this case knowledge cannot be created by data mining – due to an intrinsic lack of data in the beginning of any investigation. However, experience, human ideas and imagination lead to efficient assumptions about the tax betrayer’s tricks and typical behavior.
The tax fraud investigation can be embedded into the Bayesian Learning Theory. This approach is based on hints, investigation (unscrambling information) and integration of partial information in a stepwise procedure. The kick-off is an initial suspicion issued by an insider like a fired employee, disappointed companion or wife, envious neighbor or inquisitive custom collector. This first step can be conceived as the fixing of the prior distribution $p(\theta)$ on the tax liability size θ of the tax betrayer as an initial suspicion. The next step at the tax authority’s site is concerned with opening a new case, and getting access to the tax file of the suspect. Thereby new evidence (x) is created and hints given. Formally, the likelihood of the tax fraud, $l(x \mid \theta)$, is established. This

allows updating of the initial suspicion for gaining the posterior distribution $p(\theta \mid x) \propto l(x \mid \theta)\, p(\theta)$.
Iteration ("Learning") is performed if further step by step investigations deliver more information on the non-conforming suspect's life style related to the series of his annual taxable income and assets. The necessary investigations are tricky for getting insight into the betrayer's life style, and make use of criminal investigator's good practice like, for instance, "Simple issues first".
More formally, we take the former posterior $p(\theta \mid x)$ as a new prior $p^*(\theta)$ and combine it with the new facts (x') about the tax crime using the likelihood $l^*(x' \mid \theta)$. This leads to the updated suspicion $p^*(\theta \mid x')$ as the new posterior. The investigation stops when the tax liability is fixed and p* as a measure of certainty is near 100% ("most probably, if not certainly"). Alternatively, the tax authorities may stop it when p* drops down extremely, i.e. doubts increase. In the first case the charge is left to the judicial system to prosecute, judge and eventually arrest the betrayer.
Alternative approaches are Case Based Reasoning and Rule-based Systems and there interaction with the Bayesian approach is mentioned.
Finally, innocent and lawful people can be hopeful because betrayers never will be able to construct a perfect manipulated ("artificial") world of figures, and in the long run they will be captured as F. Wehrheim (2011) clearly pointed out.

Can We Use the Highest Reported Age at Death as Proxies of the Maximum Life Span?

Jean Marie Robine
INSERM/EPHE and INED, France

In this paper we will explore whether we can use the highest reported ages at death (HRAD), including the maximum reported age at death (MRAD), as proxies of the maximum life span (MLS). MLS is an established biological concept indicating the maximum duration of life that an individual of a specific species can expect to reach taking into account inherent biological constraints. Several values ranging from 110 to 125 years have been proposed for the human species. Highest or maximum ages at death are empirical observations.
In this paper, we will present and discuss four empirical approaches:

1. The records approach: using only the age at death of the successive oldest living persons. We hypothesize that this approach, which can provide less than one observation by year, gives excessive weight to outlier values. In some variants of this approach we will fix some age threshold.

2. The MRAD approach: using the annual maximum reported age at death (MRAD), providing one observation by year. A variant is to consider MRAD for males and females separately, doubling the number of observation.
3. The supercentenarian approach: using all death above the age of 110, providing many observations for the recent years, most of them being concentrated near 2015. This data set can be summarized by the annual mean age at death above the age of 110. This series should strongly limit the weight of outlier values.
4. The HRAD approach: using several series of high reported age at death (HRAD) such as the highest RAD, the 2nd highest RAD, the 3rd highest RAD, … the 10th highest RAD. The first series (equivalent to the MRAD series) will possibly grasp several outliers. The second series may still grasp one or two outliers but when using the 5th, 6th ETC. highest RAD series, the probability to grasp outliers should be very low.

We hypothesize that the 3rd and the 4th approaches can help disentangling between trends and noise (outliers). Combining all approaches can help discussing among the empirical proxies of MLS which empirical proxies can be used as "better" indicator of the maximum life span. Some "noisy" proxies can suggest the presence of a MLS while "less exposed" proxies can possibly suggest an ongoing increase in longevity over time.

Invited and Contributed Talks

A Topological Discriminant Analysis

Rafik Abdesselam

COACTIS-ISH Management Sciences Laboratory - Human Sciences Institute, University of Lyon, France

In this paper, we propose a new discriminant approach, called Topological Discriminant Analysis, which use a proximity measure in a topological context. The results of any operation of clustering or classification of objects strongly depend on the proximity measure chosen. The user has to select one measure among many existing ones. Yet, from a discrimination point of view, according to the notion of topological equivalence chosen, some measures are more or less equivalent. The concept of topological equivalence uses the basic notion of local neighborhood.
In a discrimination context, we first define the topological equivalence between the chosen proximity measure and the perfect discrimination measure adapted to the data considered, through the adjacency matrix induced by each measure, then propose a new topological method of discrimination using this selected proximity measure. To judge the quality of discrimination, in addition to the classical percentage of objects well classified, we define a criterion for topological equivalence of discrimination.
The principle of the proposed approach is illustrated using a real data set with conventional proximity measures of literature for quantitative variables. The results of the proposed Topological Discriminant Analysis, associated to the \best" discriminating proximity measure, are compared with those of classical metric models of discrimination, Linear Discriminant Analysis and Multinomial Logistic Regression.
Keywords: Proximity measure; Topological structure; Neighborhood graph; Adjacency matrix; Topological equivalence; discrimination.

About First-passage Times of Integrated Gauss-Markov Processes

Mario Abundo

Dipartimento di Matematica, Università Tor Vergata, Italy

First-passage time (FPT) problems for integrated Markov processes arise both in theoretical and applied Probability. In certain stochastic

models for the movement of a particle, its velocity is modeled as Brownian Motion $B(t)$ (BM), or more generally as a diffusion process $Y(t)$. Thus, particle position turns out to be the integral of $Y(t)$, and any question about the time at which the particle first reaches a given place leads to the FPT of integrated $Y(t)$. This investigation is complicated by the fact that the integral of a Markov process is no longer Markovian; however, the two-dimensional process $(\int_0^t Y(s)ds, Y(t))$ is Markovian, so the FPT of integrated $Y(t)$ can be studied by using Kolmogorov's equations. The study of $\int_0^t Y(s)ds$ has interesting applications in Biology, e.g. in the framework of diffusion models for neural activity; if one identifies $Y(t)$ with the neuron voltage at time t, then $\int_0^t Y(s)ds$ represents the time average of the neural voltage in the interval $[0,t]$. Another application can be found in Queueing Theory, if $Y(t)$ represents the length of a queue at time t; then $\int_0^t Y(s)ds$ represents the cumulative waiting time experienced by all the 'users' till the time t. We extend to integrated Gauss-Markov (GM) processes some known results about the FPT of integrated BM. We consider a continuous GM process Y of the form:

$$Y(t) = m(t) + h_2(t)\, B(\rho(t)),\ \ t \geq 0$$

where $B(t)$ is a standard BM, $m(t) = E(Y(t))$ and the covariance $c(s,t) = E\,[(Y(s)-m(s))(Y(t)-m(t))],\ 0 \leq s < t,$ are continuous functions; moreover, $c(s,t) = h_1(s)\, h_2(t)$, $\rho(t) = h_1(t)/\, h_2(t)$ is a non-negative, monotonically increasing function with $\rho(0) = 0$. Besides BM, a noteworthy case of GM process is the Ornstein-Uhlenbeck (OU) process. Given a GM process Y, consider the integrated process, starting from $X(0) = x$, i.e. $X(t) = x + \int_0^t Y(s)ds$; for a given boundary $a > x$, we study the FPT of X through α, with the conditions that $X(0) = x$ and $Y(0) = m(0) = y$, that is:

$$\tau_a\,(x,y) = \inf\,\{t > 0 \colon X(t) = a \mid X(0) = x,\ Y(0) = y\},$$

as well as, for $x \in (a,b)$, the first-exit time of X from the interval (a,b), with the conditions that $X(0) = x$ and $Y(0) = y$, that is:

$$\tau_{ab}\,(x,y) = \inf\,\{t > 0 \colon X(t) \notin (a,b) \mid X(0) = x,\ Y(0) = y\}.$$

An essential role is played by the representation of X in terms of BM, obtained by us in (Abundo, 2013). By using the properties of continuous martingales, we reduce the problem to the FPT of a time-changed BM. In the one-boundary case, we present an explicit formula for the density of the FPT, while in the two-boundary case, we are able to express the n-th order moment of the first-exit time as a series involving only elementary functions.

References

1. Abundo, M., 'On the representation of an integrated Gauss-Markov process'. Scientiae Mathematicae Japonicae Online, e-2013, 719–723.

The IPUMS Database in the Estimation of Infant Mortality Worldwide

Alejandro Aguirre, Fortino Vela Peón
El Colegio de México, México

William Brass (Brass et al., 1968) developed the indirect method of children ever born /children surviving (CEB/CS) to estimate infant and child mortality. The CEB/CS method uses information (usually from censuses, although it may come from surveys) of the total number of children ever born (CEB), and the children surviving (CS) that women have had throughout their life, until the moment in which they are interviewed. The information is classified by age of the mother. It is expected (on average) that the older the women, the higher the risk of death for their children, because they have been exposed to the risk of death during a longer period, and thus the proportion of dead children increases with the age of the woman.
The Integrated Public Use Microdata Series (IPUMS) is a project of the University of Minnesota that basically consists on collecting and distributing census data from all over the world. Within its goals are to collect and preserve data and documentation, as well as to harmonize the data. They have gathered 277 censuses from 82 countries.
In many censuses the questions on CEB/CS have been asked, mainly since the second half of the XX century. In this paper we estimate infant mortality, for all the censuses available in the IPUMS database that contain the necessary information. We contrast these results with those obtained using vital statistics.
Keywords: infant mortality; indirect estimation; Brass; Integrated Public Use Microdata Series (IPUMS)

Characterizations of Distributions by Extended Samples

M. Ahsanullah
Rider University, USA

Suppose we have m observations from an absolutely continuous distribution. We order these observations in increasing order. We take another n-m (n>m) observations from the same distribution and order them. We consideration characterization of distribution based on the jth of observation of m samples based of ith sample of n samples. Several new results are presented.
Keywords: Characterization, Order Statistics, Exponential Distribution, Pareto distribution and Uniform distribution.

Mortality Modelling Using Probability Distributions: Application in Greek Mortality Data

Andreopoulos Panagiotis[1], Bersimis G. Fragkiskos[2], Tragaki Alexandra[1], Rovolis Antonis[3]

[1]Dept of Geography, Harokopio University, Greece,
[2]Dept of Informatics and Telematics, Harokopio University, Greece,
[3]Dept of Geography, Harokopio University, Greece,
[3]Dept of Economic and Regional Development, Greece

A number of different distributions describing age-related mortality have been proposed. The most common ones, Gompertz and Gompertz - Makeham distributions have received wide acceptance and describe fairly well mortality data over a period of 60-70 years, but generally do not give the desired results for old and/or young ages. This paper proposes a new mathematical model, combining the above distributions with Beta distribution. Beta distribution was chosen for its flexibility on age-specific mortality characteristics. The proposed model is evaluated for its goodness of fit and showed sufficient predictive ability for different population sub-groups. The scope of this work is to create sufficient mortality models that could also be applied in populations other than the Greek, based on appropriate parameter detection (e.g. Maximum Likelihood). An examination for possible differences in the parameters' values, of the proposed model, between sexes and geographical regions (North vs South) was also attempted. The application relies on mortality data collected and provided by the Hellenic Statistic Authorities for year 2011. Population data were used in order to calculate age and sex-specific mortality rates based on the estimated mean population of one-year interval age-group for the year concerned. According to our initial findings, the proposed mortality model (ANBE) presents satisfactory results on appropriate evaluation criteria (AIC, BIC). This paper presents some of the statistical properties of the ANBE model.
Keywords: Beta distribution, Generalized Gompertz Makeham, Mortality, Spatial Analysis.

Fractal Analysis of Chaotic Processes in Living Organisms

Valery Antonov[1], Artem Zagainov[1], Anatoly Kovalenko[2],

[1]Peter the Great Saint-Petersburg Polytechnic University, Russia,
[2]Ioffe Physical-Technical Institute of Russian Academy of Sciences, Russia

The work is a generalization of the results of research conducted by the authors for several years. The main area of research was the

development and introduction of modern methods of diagnostic of the body state in real time. To create methods for rapid diagnosis of the condition of the body is designed and put into practice hardware and software package. To do this, the method of chaotic dynamics and fractal analysis of the electrocardiogram based on calculating the correlation entropy has been applied. The results of processing biological signals are shown in the form of graphs and photographs monitor. The results, which can be regarded as a system of support for operational decision-making, have shown high efficiency analysis of the body, including the transition to a critical state.
Keywords: chaotic processes, fractal analysis, body state

An Introduction to DataSHIELD

Demetris Avraam
School of Social and Community Medicine, University of Bristol, UK

Research in modern biomedicine and social science is increasingly dependent on the analysis and interpretation of individual-level data (microdata) or on the co-analysis of such data from several studies simultaneously. However, sharing and combining individual-level data is often prohibited by ethico-legal constraints and other barriers such as the control maintenance and the huge samples sizes. DataSHIELD (Data Aggregation Through Anonymous Summary-statistics from Harmonised Individual-levEL Databases) provides a novel tool that circumvents these challenges and permits the analysis of microdata that cannot physically be pooled. This presentation is an overview to DataSHIELD introducing this approach and discussing its challenges and opportunities.
Keywords: sensitive data, data analysis, data aggregation, software

A Mechanistic Model of Mortality Dynamics

Demetris Avraam, Bakhtier Vasiev
Dept of Mathematical Sciences, University of Liverpool, UK

Mortality rate in human populations increases exponentially with age in a wide range of lifespan satisfying the Gompertz law. The exponential function describing the Gompertz law occurs naturally as a mathematical solution of the equation $d\mu/dx=\beta\mu$. This equation is based on an assumption that the rate of change of the force of mortality is proportional to the force of mortality. Besides the observation that age-specific mortality increases exponentially, some deviations from this

increase exist at young and extremely old ages. A model that considers the heterogeneity of populations and expresses the force of mortality as a mixture of exponential terms has been recently shown to precisely reproduce the observed mortality patterns and explain their peculiarities at early and late life intervals. In this work, assuming that age-specific mortality data can be represented as a mixture of exponential functions, we develop a mechanistic model of mortality dynamics based on a system of linear differential equations where its solution is expressed as a superposition of exponents. The variables of the differential equations describe physiological and biological processes that affect the mortality rates. In particular, mortality data for intrinsic causes of death can appear as a solution of a coupled system of two differential equations (superposition of two exponents). The two variables in the model should be associated with the physiological states (i.e. vulnerability to diseases and ability to recover) of each individual in a population. Such model can easily be fit to mortality data for intrinsic causes of death and even extended to reproduce the total mortality dynamics. The extrinsic (mainly accidental) mortality can be modelled by a stochastic process (i.e. including an extra term on the mechanistic model described by a Poisson process).

Keywords: mortality rates, demography, mathematical modelling

Bayesian Modelling of Temperature Related Mortality with Latent Functional Relationships

Robert G Aykroyd

Dept of Statistics, University of Leeds, UK

It is common for the mortality rate to increase during periods of extreme temperature, producing a U or J-shaped mortality curve, and for the minimum mortality rate and the corresponding temperature to depend on factors such as the mean summer temperature. Previous analyses have considered long time series of temperature and mortality rate, and other demographic variables, but have ignored spatial structure. In this paper, local correlation is explicitly described using a generalized additive model with a spatial component which allows information from neighbouring locations to be combined. Random walk and random field models are proposed to describe temporal and spatial correlation structure, and MCMC methods used for parameter estimation, and more generally for posterior inference. This makes use of existing data more efficiently and will reduce prediction variability. The methods are illustrated using simulated data based on real mortality and temperature data.

Keywords: Bayesian methods, demography, generalised additive models, maximum likelihood, spatial, temporal.

Shapes Classification by Integrating Currents and Functional Data Analysis

S. Barahona[1], P. Centella[1], X. Gual-Arnau[2], M.V. Ibáñez[3], A. Simó[3]

[1]Dept of Mathematics, Universitat Jaume I. Spain,
[2]Dept of Mathematics-INIT, Universitat Jaume I., Spain,
[3]Dept of Mathematics-IMAC, Universitat Jaume I., Spain

Shape classification is of key importance in many scientific fields. This work is focused on the case where a shape is characterized by a current. A current is a mathematical object which has been proved relevant to model geometrical data, like submanifols, through integration of vector fields along them. As a consequence of the choice of a vector-valued Reproducing Kernel Hilbert Space (RKHS) as a test space to integrating manifolds, it is possible to consider that shapes are embedded in this Hilbert Space. A vector-valued RKHS is a Hilbert space of vector fields similar to $\mathbb{R}^n$, therefore it is possible to compute a mean of shapes, or to calculate a distance between two manifolds. This embedding enables us to consider classification algorithms of shapes.
We describe a method to apply standard Functional Discriminant Analysis in a vector-valued RKHS. In this, an orthonormal basis is searched by using eigenfunctions decomposition of the Kernel. This representation of data allows us to obtain a finite-dimensional representation of our sample data and to use standard Functional Data Analysis in this space of mappings. The main contribution of this method is to apply the theory of vector-valued RKHS by using currents to represent manifolds in the context of functional data.
Keywords: Currents, Statistical Shape Analysis, Reproducing Kernel Hilbert Space, Functional Data Analysis, Discriminant Analysis.

An Entropic Approach to Assess Scientific Research Output

Luiza Bădin[1,2], Anca Şerban Oprescu[1], Silvia Dedu[1], Florentin Şerban[1,3]

[1]Bucharest University of Economic Studies, Romania, [2]Gh. Mihoc - C. Iacob Institute of Mathematical Statistics and Applied Mathematics, Romania,
[3]University of Bucharest, Romania

When an organization is undertaking a development strategy in some field, it will usually need to strike a balance between the various elements that make up the overall development strategy. It is therefore important to be able to assign rankings to these elements. Usually, the elements which comprise the overall strategy will differ considerably in their character. Very few empirical studies have been conducted regarding the quantitative evaluation of entities that have quite different characters. Entropy provides a way of addressing this problem in real world situations. We also propose an algorithm for computing the weights of different indices, which allows evaluating the degree of importance of each criterion considered in the analysis. Computational results are provided.

It is important to note that our algorithm can be used with various types of entropy measures. In the future it would be important to try to establish which entropy measure should be used on the data set, in order to provide real world conclusions, or as close as is possible to this. The aim of future research would therefore be to address this issue and to improve the fit between the model and reality.

Keywords: research-development activity, entropy, rankings, weights, comprehensive assessment, standardization.

Acknowledgment: *This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2905.*

Variance Reduction of the Mean Number of Customers in the System of M/M/1 Retrial Queues Using RDS Method

Latifa BAGHDALI-OURBIH, Khelidja IDJIS, Megdouda OURBIH-TARI

ENSTP, Algeria

Retrial queues have been widely used to model many problems in telephone switching systems, telecommunication networks, computer networks and computer systems. Various techniques and results have been developed, either to solve particular problems or to understand the basic stochastic processes. The advances in this area are summarized in review articles of Yang and Templeton (1987) and Falin (1990). In simulation, the standard sampling procedure used to represent the stochastic behavior of the input random variables is Simple Random Sampling (SRS), the so-called Monte Carlo (MC) method (Dimov, 2008, Robert and Casella, 2004). This method is well known and used in an intensive way, it can solve a large variety of problems, but statistically it is not the best, because its estimates obtained through simulation vary between different runs. As a consequence, other sampling methods were proposed to reduce the variance of MC estimates. We can cite

Refined Descriptive Sampling (RDS) (Tari and Dahmani, 2006). This method was proposed as a better approach to MC simulation, so, it is used to be compared to SRS.
This paper simulates the stationary M/M/1 retrial queues using SRS and RDS to generate input variables. We design and realize software under Linux using C language which establishes the number of customers in the system of the M/M/1 retrial queues, computes the relative deviation and the variance reduction in order to compare both sampling methods. The mean number of customers in the system was estimated using both sampling methods. The studied performance measure of the M/M/1 retrial queue is given by (Falin and Templeton, 1997).
The simulation results demonstrate that RDS produces more accurate and efficient point estimates of the true parameter and can significantly improves the mean number of customers in the system sometimes by an important variance reduction factor in the M/M/1 retrial queue compared to SRS.
Keywords: Simulation; Retrial Queues; Sampling; Monte Carlo; Variance reduction.

Germination and Seedling Emergence Model of Bush Bean and Maize in Different Physical Soil Characteristics

Behnam Behtari, Adel Dabbag Mohammadi Nasab
Dept of Crop Ecology, Uinversity of Tabriz, East Azarbaijan, Iran

A field study was carried out to investigation of effects of four depth planting and three soil types with different physical characteristics on bush bean (Phaseolus vulgaris var. sunray) and maize (Zea mays L. var. Amyla) seed germination and seedling emergence. The aim of the experiments was to investigate if the physical characteristics of the soils were involved in both buried seed ecology and emergence dynamics. The result revealed that germination inhibition due to burial depth was found to be directly proportional to clay content and inversely proportional to sand content. Depth of fifty percent emergence inhibition (Di50%) in clay soil for both bush bean and maize were equal to 5.3 cm, if this was for silty soil respectively 5.4 and 2.7 cm. Significant ($p < 0.01$) linear regressions between clay particle content and Di50% revealed that those soil component had opposite effects in terms of favoring or inh! ibiting depth mediated inhibition. Therefore, with increasing the amount of clay soil, the amount of inhibition increases. The data obtained from these experiments show that the oxygen content in the surrounding soil of seeds can not be an important factor for seed germination differences, so that the effect it was not significant. With increasing geometric mean particle diameter soil inhibition decreased. In conclusion, these

experiments demonstrated that soil physical properties have a strong effect on buried-seed ecology and consequently on seed germination and seedling emergence.

Keywords: Depth of 50% emergence inhibition, Geometric mean particle diameter, Soil clay, Soil texture

Weighting as a Method of Optimizing an Index's Diagnostic Performance: The Case of Attica Study in Greece

Fragkiskos G. Bersimis[1], Demosthenes Panagiotakos[2], Malvina Vamvakari[1]

[1]Dept of Informatics and Telematics, Harokopio University, Greece,
[2]Dept of Nutrition Science - Dietetics, Harokopio University, Greece

In this work, the use of weights in a composite health related index, which is constructed by *m* variables, is evaluated whether differentiates its diagnostic accuracy. An un-weighted index and weighted indices were developed by using various weighting methods. The health related indices' diagnostic ability was evaluated by using suitable measures, such as, sensitivity, specificity and AUC. In this work, logistic regression and discriminant analysis were chosen as distinguishing methods, between patients and healthy individuals, for generating corresponding weights. Values of AUC, sensitivity and specificity of weighted indices were significantly higher compared to the un-weighted one. The theoretical results were applied in dietary data collected from ATTIKA research in Greece. In addition, weighted indices were more effective in classifying individuals from ATTIKA research as patients or non-patients correctly compared to the unweighted index.

Keywords: Health index; ROC; Discriminant analysis; Logistic regression; Sensitivity; Specificity.

References

1. Streiner, D.L., Norman, G.F. (2008) Introduction. In Health Measurement Scales, 4th Ed.; Oxford University Press, USA, pp. 1-4.
2. Panagiotakos D. (2009) Health measurement scales: methodological issues. Open Cardiovasc. Med. J.3, pp. 160-5
3. Bersimis, F., Panagiotakos, D., Vamvakari, M.: Sensitivity of health related indices is a non-decreasing function of their partitions. Journal of Statistics Applications & Probability 2(3), 183–194 (2013)

A Sequential Test for Assessing Agreement between Raters

Sotiris Bersimis[1], Athanasios Sachlas[1], Subha Chakraborti[2]

[1]Dept of Statistics & Insurance Science, University of Piraeus, Greece,
[2]Dept of Information Systems, Statistics and Management Science,
University of Alabama, USA

Assessing the agreement between two or more raters is an important aspect in medical practice. Existing techniques, which deal with categorical data, are based on contingency tables. This is often an obstacle in practice as we have to wait for a long time to collect the appropriate sample size of subjects to construct the contingency table. In this paper, we introduce a nonparametric sequential test for assessing agreement, which can be applied as data accrues, does not require a contingency table, facilitating a rapid assessment of the agreement. The proposed test is based on the cumulative sum of the number of disagreements between the two raters and a suitable statistic representing the waiting time until the cumulative sum exceeds a predefined threshold. We treat the cases of testing two raters' agreement with respect to one or more characteristics and using two or more classification categories, the case where the two raters extremely disagree, and finally the case of testing more than two raters' agreement. The numerical investigation shows that the proposed test has excellent performance. Compared to the existing methods, the proposed method appears to require significantly smaller sample size with equivalent power. Moreover, the proposed method is easily generalizable and brings the problem of assessing the agreement between two or more raters and one or more characteristics under a unified framework, thus providing an easy to use tool to medical practitioners.
Keywords: Agreement assessment, Cohen's k, Hypothesis testing, Markov Chain embedding technique, Reliability, Sequential testing.

Dependent Credit-Rating Migrations: Coupling Schemes, Estimators and Simulation

Dmitri V. Boreiko[1], Serguei Y. Kaniovski[2], Yuri M. Kaniovski[1], Georg Ch. Pflug[3]

[1]Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy, [2]Austrian Institute for Economic Research (WIFO), Austria,
[3]Dept of Statistics and Decision Support Systems, University of Vienna, Austria

By mixing an idiosyncratic component with a common one, coupling schemes allow to model dependent credit-rating migrations. The distribution of the common component is modified according to macroeconomic conditions, favorable or adverse, that are encoded by the corresponding (unobserved) tendency variables as 1 and 0. Computational resources required for estimation of such mixtures

depend upon the pattern of tendency variables. Unlike in the known coupling schemes, the credit-class-specific tendency variables considered here can evolve as a (hidden) time-homogeneous Markov chain. In order to identify unknown parameters of the corresponding mixtures of multinomial distributions, maximum likelihood estimators are suggested and tested on Standard and Poor's dataset using MATLAB optimization software.

Keywords: coupled Markov chain, mixture, (hidden) Markov chain, maximum likelihood, default, Monte-Carlo simulation.

Joint Modelling of Longitudinal Tumour Marker CEA Progression and Survival Data on Breast Cancer

Ana Borges, Inês Sousa, Luis Castro

Porto Polytechnic Institute - School of Management and Technology of Felgueiras (ESTGF - IPP), Center for Research and Innovation in Business Sciences and Information Systems (CIICESI), Portugal

The work proposes the use of statistical methods within the biostatistics to study breast cancer in patients of Braga's Hospital Senology Unit. With the primary motivation to contribute to the understanding of the progression of breast cancer, within the Portuguese population, using a more complex statistical model assumptions than the traditional analysis. The analysis preformed has as main objective to develop a joint model for longitudinal data (repeated measurements over time of a tumour marker) and survival (time-to-event of interest) of patients with breast cancer, being death from breast cancer the event of interest. The data analysed gathers information on 540 patients, englobing 50 variables, collected from medical records of the Hospital. We conducted a previous independent survival analysis in order to understand what the possible risk factors for death from breast cancer for these patients. Followed by an independent longitudinal analysis of tumour marker Carcinoembryonic antigen (CEA), to identify risk factors related to the increase in its values. For survival analysis we made use of the Cox proportional hazards model (Cox, 1972) and the flexible parametric model Royston-Parmar (Royston & Parmar, 2002). Generalized linear mixed effect models were applied to study the longitudinal progression of the tumour marker. After the independent survival and longitudinal analysis, we took into account the expected association between the progression of the tumour marker values with patient's survival, and as such, we proceeded with a joint modelling of these two processes to infer on the association between them, adopting the methodology of random effects. Results indicate that the longitudinal progression of CEA is significantly associated with the probability of survival of these patients. We also conclude that as the independent analysis returns

biased estimates of the parameters, it is necessary to consider the relationship between the two processes when analysing breast cancer data.

Keywords: Joint Modelling, survival analysis, longitudinal analysis, Cox, random effects, CEA, Breast Cancer

Applications of the Cumulative Rate to Kidney Cancer Statistics in Australia

Janelle Brennan[1], K.C. Chan[2], Rebecca Kippen[3], C.T. Lenard[4], T.M. Mills[5], Ruth F.G. Williams[4]

[1]Dept of Urology, Bendigo Health, and St. Vincent's Hospital Melbourne, Australia, [2]Computer Science and Information Technology, La Trobe University, Bendigo, Australia, [3]School of Rural Health, Monash University, Australia, [4]Mathematics and Statistics, La Trobe University, Australia, [5]Bendigo Health and La Trobe University, Bendigo, Australia

Cancer incidence and mortality statistics in two populations are usually compared by use of either age-standardised rates or cumulative risk by a certain age. We argue that the cumulative rate is a superior measure because it obviates the need for a standard population, and is not open to misinterpretation as is the case for cumulative risk. Then we illustrate the application of the cumulative rate by analysing incidence and mortality data for kidney cancer in Australia using the cumulative rate. Kidney cancer is also known as malignant neoplasm of kidney: we use the term kidney cancer in this paper. We define kidney cancer as the disease classified as C64 according to the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD10) by Australian Institute of Health and Welfare. Kidney cancer is one of the less common cancers in Australia. In 2012, approximately 2.5% of all new cases of cancer were kidney cancer, and approximately 2.1% of all cancer related deaths in Australia are due to kidney cancer. There is variation in incidence and mortality by sex, age, and geographical location in Australia. We examine how the cumulative rate performs in measuring the variation of this disease across such subpopulations. This is part of our effort to promote the use of the cumulative rate as an alternative to the age-standardised rates or cumulative risk. In addition, we hope that this statistical investigation will contribute to the aetiology of the disease from an Australian perspective.

Keywords: Kidney cancer, Incidence, Mortality, Cumulative rate, Descriptive epidemiology.

The Contribution of Non-Communicable Disease to the Burden of Deaths in Rural Burkina Faso

Bunker A[1,2], Marquins S[3], Sié A[4], Rocklov J[3], Sauerborn R[2]

[1]Network Aging Research, University of Heidelberg, Germany, [2]Public Health Institute, Heidelberg University, Heidelberg, Germany, [3]Dept of Public Health and Clinical Medicine, Epidemiology and Global Health, Umeå University, Sweden, [4] Centre de Recherche en Santé de Nouna, Burkina Faso

We investigated the contribution of non-communicable disease (NCD) to the total burden of premature deaths in the years 2000-2010 using data collected by the Nouna Health and Demographic Surveillance System (HDSS) in rural Burkina Faso. The Nouna HDSS is part of the International Network for the Demographic Evaluation of Populations and their Health (INDEPTH), which collects health and demographic data across populations in low-and-middle income countries.

A life table was constructed for the Nouna HDSS. Female life expectancy at birth was higher; 59.7 years in comparison to male life expectancy; 56.8 years. We calculated all-cause, communicable, non-communicable and injury-related years of life lost (YLL) as an indicator of premature mortality for every death by matching age and gender to the life table. Apart from the 0-1 and 1-4 age group, abridged life tables were created in five-year increments to produce stable life-expectancy estimates for a relatively small population.

Of the all-cause 323168 YLLs (number of deaths (n)=8060), NCDs contributed to 22210 YLLs (n=934). Communicable disease and injuries were associated with 246337 (n=5590) and 12647 (n=309) YLLs respectively. In a society traditionally affected by communicable diseases, particularly malaria, NCDs account for about 7% of YLLs. As life expectancy increases in Burkina Faso and people live for longer with chronic and degenerative diseases, NCDs will play a greater role in contributing to the burden of premature deaths.

Key words: Years of life lost, Non-communicable disease, Sub-Saharan Africa.

The Difficulties of the Access to Credit for the SMEs: An International Comparison

Raffaella Calabrese[1], Cinzia Colapinto[2], Anna Matuszyk[3], Mariangela Zenga[4]

[1]Essex Business School, University of Essex, United Kingdom,
[2]Dept of Management, Ca' Foscari University of Venice, Italy,
[3]Institute of Finance, Warsaw School of Economics, Poland,
[4]Dept of Statistics and Quantitative methods, Milano-Bicocca University, Italia

Small and medium enterprises (SMEs) play a significant role in their economies as key generators of employment and income and as drivers of innovation and growth. This is even more important in the perspective of the economic recovery from the 2008 economic and financial crisis.
The crisis has had a negative impact on bank lending, and likely on SMEs' life as well. Indeed, in the case of reduced bank lending SMEs tend to be more vulnerable and affected than larger companies. Great attention has been paid to the situation of SMEs due to the risks of a credit crunch and increase in the financing gap in Europe.
We run a survey on access to finance for SMEs in order to gain a better understanding of access to credit by SMEs in three European countries, namely Italy, Poland and United Kingdom. The survey aims at identifying the main difficulties faced by SMEs in trying to raise credit and to understand if their financial resources are adequate. Moreover, an Indicator of Financial Suffering is built as a composite indicator. By using the data mining techniques, we are able to identify the SMEs' groups that experienced greater difficulties in accessing founds to finance the business.
Keywords: SME, Financial crisis, Access to credit, Index of Financial suffering, Data Mining techniques.

Modeling Mortality Rates using GEE Models

Liberato Camilleri[1], Kathleen England[2]

[1]Dept of Statistics and Operations Research, University of Malta, Malta,
[2]Directorate of Health Information and Research, Malta

Generalised estimating equation (GEE) models are extensions of generalised linear models by relaxing the assumption of independence. These models are appropriate to analyze correlated longitudinal responses which follow any distribution that is a member of the exponential family. This model is used to relate daily mortality rate of Maltese adults aged 65 years and over with a number of predictors, including apparent temperature, season and year. To accommodate the

right skewed mortality rate distribution a Gamma distribution is assumed. An identity link function is used for ease of interpretating the parameter estimates. An autoregressive correlation structure of order 1 is used since correlations decrease as distance between observations increases. The study shows that mortality rate and temperature are related by a quadratic function. Moreover, the GEE model identifies a number of significant main and interaction effects which shed light on the effect of weather predictors on daily mortality rates.
Keywords: Generalised estimating equation, Daily mortality rates, Apparent temperature.

Numerical Methods on European Option Second Order Asymptotic Expansions for Multiscale Stochastic Volatility

Betuel Canhanga[1,2], Anatoliy Malyarenko[2], Jean-Paul Murara[2,3], Ying Ni[2], Milica Ranic[2], Sergei Silvestrov[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, University, Sweden, [3]College of Science and Technology, School of Sciences, Dept of Applied Mathematics, University of Rwanda, Rwanda

State of art, after Black-Scholes proposed in 1973 a model for pricing European Options under constant volatilities, Christoffersen in 2009 empirically showed "why multi factor stochastic volatility models work so well". Four years late Chiarella and Ziveyi solved the Christoffersen model considering an underlying asset whose price is governed by two factor stochastic volatilities. Using the Duhamel's principle, they derived an integral form solution of the boundary value problem associated to the option price, applying method of characteristics, Fourier transforms and Laplace transforms they computed an approximate formula for pricing American options.
In this paper, considering Christoffersen model, we assume that the two factor stochastic volatilities in the model are of mean reversion type one changing fast and another changing slowly. Continuing our previous research where we provided experimental and numerical studies on investigating the accuracy of the approximation formulae given by the first order asymptotic expansion, here we present experimental and numerical studies for the second order asymptotic expansion and we compare the obtained results with results presented by Chiarella and Ziveyi and also with the results provided by the first order asymptotic expansion.
Keywords: Option pricing model, asymptotic expansion, numerical studies.

On Stepwise Increasing Roots of Transition Matrices

Philippe Carette[1], Marie-Anne Guerry[2]

[1]*Dept of General Economics, University of Ghent, Belgium,*
[2]*Dept Business Technology and Operations, Vrije Universiteit Brussel, Belgium*

In Markov chain models, given an empirically observed transition matrix over a certain time interval, it may be needed to extract information about the transition probabilities over some shorter time interval. This is called an embedding problem (see e.g. [1], [3]). In a discrete time setting this problem comes down to finding a transition matrix Q which is a stochastic p-th root (p is an integer) of a given transition matrix P. It is known that an embedding problem need not have a unique solution ([2]), so the question arises as to identify those solutions that can be retained for further modelling purposes ([3]). In manpower planning applications, it is reasonable to assume that promotion prospects decrease over shorter periods of time ([4]). Therefore, we focus on transition matrices Q which have off-diagonal elements that are not exceeding the corresponding elements of P and call those matrices *stepwise increasing.*

In this paper, we present some results about stepwise increasing stochastic square roots (p = 2) of a given transition matrix for the two- and three-state case.

Keywords: Markov chain, embedding problem, transition matrix, identification problem.

References

1. B. Singer and S. Spilerman. The representation of social processes by Markov models. American Journal of Sociology, 1-54, 1976.

2. N. J. Higham and L. Lin. On pth roots of stochastic matrices. Linear Algebra and its Applications, 435, 3, 448-463, 2011.

3. R. B. Israel, J. S. Rosenthal and J. Z. Wei. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. Mathematical finance, 11, 2, 245-265, 2001.

4. M.A. Guerry. Some Results on the Embeddable Problem for Discrete-Time Markov Models in Manpower Planning, Communications in Statistics-Theory and Methods, 43, 7, 1575-1584, 2014.

Predicting and Correlating the Strength Properties of Wood Composite Process Parameters by Use of Boosted Regression Tree Models

Dillon M. Carty, Timothy M. Young, Frank M. Guess, Alexander Petutschnigg

University of Tennessee, USA

Predictive boosted regression tree (BRT) models were developed to predict modulus of rupture (MOR) and internal bond (IB) for a US particleboard manufacturer. The temporal process data consisted of 4,307 records and spanned the time frame from March 2009 to June 2010. This study builds on previous published research by developing BRT models across all product types of MOR and IB produced by the particleboard manufacturer. A total of 189 continuous variables from the process line were used as possible predictor variables. BRT model comparisons were made using the root mean squared error for prediction (RMSEP) and the RMSEP relative to the mean of the response variable as a percent (RMSEP%) for the validation data sets. For MOR, RMSEP values ranged from 1.051 to 1.443 MPa, and RMSEP% values ranged from 8.5 to 11.6 percent. For IB, RMSEP values ranged from 0.074 to 0.108 MPa, and RMSEP% values ranged from 12.7 to 18.6 percent. BRT models for MOR and IB predicted better than respective regression tree models without boosting. For MOR, key predictors in the BRT models were related to “pressing temperature zones,” “thickness of pressing,” and “pressing pressure.” For IB, key predictors in the BRT models were related to “thickness of pressing.” The BRT predictive models offer manufacturers an opportunity to improve the understanding of processes and be more predictive in the outcomes of product quality attributes. This may help manufacturers reduce rework and scrap and also improve production efficiencies by avoiding unnecessarily high operating targets.

Keywords: Regression Trees, Boosted Regression Trees, Predictive Modeling, Modulus

Nonparametric Estimation of the measure associated with the Lévy-Khintchine Canonical Representation

Mark Anthony Caruana

Dept of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

Given a Lévy process $\{X_t\}_{t\geq 0}$ observed on a finite time interval $[0,T]$, we consider the nonparametric estimation of the function H, sometimes called the jump function, associated with the Lévy-Khintchine Canonical representation over an interval $[c,d]$ where $0 < c < d$. In particular, we shall assume a high-frequency framework and apply the method of sieves to estimate *H*. We also show that under certain conditions the estimator enjoys asymptotic normality and consistency. The dimension of the sieve and the length of the estimation interval $[c,d]$ will be investigated. Finally, a number of simulations will also be conducted.

Keywords: nonparametric estimation, method of sieves, Lévy-Khintchine Canonical representation, asymptotic normality, high-frequency framework.

An Analysis of a Rubin and Tucker Estimator

Mark Anthony Caruana, Lino Sant

Dept of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

The estimation of the Lévy triple is discussed in a paper by Rubin and Tucker. Although this work is mentioned in numerous papers related to nonparametric inference of Lévy processes, these estimators are rarely ever used or implemented on data sets. In this paper we shall study some properties of the estimator of the Lévy measure which features in the paper of aforementioned authors, outlining its strengths and weaknesses. In particular, we consider some convergence rates. Finally, a number of simulations are presented and discussed.

Keywords: Lévy Triple, Lévy Measure, convergence rates.

Measuring Inequality in Society

K.C. Chan[1], C.T. Lenard[2], T.M. Mills[3], Ruth F.G. Williams[2]

[1]*Computer Science and Information Technology, La Trobe University, Australia,*
[2]*Mathematics and Statistics, La Trobe University, Australia,*
[3]*Bendigo Health and La Trobe University, Australia*

Some inequalities in society seem unfair. Hence, governments around the world develop policies aimed at reducing unwarranted inequalities. To assess the impact of these policies, one needs well-founded measures of inequality to monitor the impact of such policies. Are these policies effective in moving society to be more egalitarian or not? The

discipline of economics has developed many such measures over the last century: examples include the Lorenz curve, the measures of Gini, Atkinson and Theil. Although these measures have been focussed on measuring inequality of incomes, they have been adapted to other situations too. In this expository paper, we will present an introduction to measures of inequality in society from a mathematical perspective, and highlight the variety of applications. This is an area in which mathematics can contribute to justice and fairness in society.
Keywords: Inequality, Economics, Lorenz, Gini, Atkinson, Theil, Axioms

Spatial Bayes Lung Cancer Incidence Modelling

Janet Chetcuti, Lino Sant

Dept of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

Spatio-temporal variability maps of disease incidence rates are very useful in establishing effective risk factors and determining their relative importance. Assembling disease mappings involves going back in time and capturing variations by small areas. National medical records are not always as detailed, geographically extensive and freely available as one would need. Records over the period 1995-2012 for lung cancer incidence in the Maltese Islands are a case in point. They constitute the point of departure of this paper in which modelling techniques have been suitably selected to provide appropriate disease maps. Resulting models have to take into account only effects and factors for which information can indeed be extracted from the data available. In particular Bayesian techniques offer a powerful repertoire of models which provide the stochastic basis and mathematical validity for capturing expressive spatial and temporal interactions underwritten by a system of consistent, conditional probabilities. Conditional autoregressive models were developed and estimated in a way which enabled geographical considerations to be incorporated into a correlation structure, shared by a hierarchical pyramid of random variables, out of which a random field can be defined.
The Bayesian models, computed with the use of MCMC algorithms, should help establish significant links between pockets of rates from different Maltese localities.
Keywords: lung cancer incidence, random fields, hierarchical Bayesian spatio-temporal models.

Almost Graduated, Close to Employment? Taking into Account the Characteristics of Companies Recruiting at a University Job Placement Office

Franca Crippa, Mariangela Zenga, Paolo Mariani

Dept of Phychology, University of Milano-Bicocca, Italy

Higher education employability is a major concern in recent years, in terms of the success in finding a job, possibly a 'good job' (Clark, 1998), afer graduation. Since Italian universities became intermediaries between their graduates and the labour market, according to the law 30/2003, their Job Placement offices have played a key role in the interplay between students at the end of the university tracks and companies in need of fresh professionals, so as to fulfill higher education's mission in its last phase, the entry in the labour market. Present academic data sources therefore provide several elements useful in understanding not only internal educational processes, but also, potentially, their links with other social actors, thereby expanding the viewpoint to some productive reality. In this respect, more than 4000 companies, registered with the Portal of Almalaurea for recruitment and linkage with the Job Placement Office of the University of Milano-Bicocca, adhered in June 2015 to an electronic statistical investigation, in the frame of a multicentre research, providing their structural characteristics together with some attitudes in defining the positions they need.

Statistical methods for the analysis of students' careers at their exit stage are hence explored, in order to grasp the reverted perspective in education, from companies searching for graduated positions directly within the pertinent educational institution. Firstly, companies' characteristics are considered, so as to understand the productive areas that currently benefit from a direct relationship with universities. As a subsequent goal, the feasibility of extending students' university paths to the stage of companies' selection, using Markov Chains with Fuzzy States (Crippa, Mazzoleni and Zenga, 2015), is explored also in simulations.

Keywords: higher education, employability, transition

Population Ageing and Demographic Aspects of Mental Diseases in the Czech Republic

Kornélia Cséfalvaiová[1], Jana Langhamrová[2], Jitka Langhamrová[1]

[1]Dept of Demography, University of Economics, Czech Republic,
[2]Dept of Statistics and Probability, University of Economics, Czech Republic

In developed societies, mainly due to a progress in healthcare and medicine, it is increasingly probable to reach old ages and survive old age. Adult mortality is decreasing, and for this reason human populations live longer. As people live longer and populations are ageing, the total expenditure on health and healthcare is also increasing and represents an important part of the government budget. There are some discrepancies among the countries, but this issue ultimately leads to higher government's spending on public health. In our study we focus on the mental health problems in the Czech Republic and selected European countries. Mental diseases are one the major public health objects in ageing societies that requires our attention.
Keywords: Alzheimer's Disease, Czech Republic, Mental Diseases, Population Ageing.

Life Annuity Portfolios: Solvency Assessing and Risk-Adjusted Valuations

Valeria D'Amato[1], Emilia Di Lorenzo[2], Albina Orlando[3], Marilena Sibillo[1]

[1]Dept of Economics and Statistics, Campus Universitario, University of Salerno, Italy, [2]Dept of Economic and Statistical Sciences, University of Naples Federico II, Italy, [3]National Research Council, Italy

Solvency assessing is a compelling issue for insurance industry, also in light of the current international risk-based regulations. Internal models have to take into account risk/profit indicators in order to provide flexible tools aimed at valuing solvency. Considering a portfolio of life annuities (as well as saving products), we deepen this topic by means of the {\em solvency ratio}, which properly captures both financial and demographic risk drivers.
The analysis is carried out in accordance with a management perspective, apt to measure the business performance, which requires a correct risk control.
In the case of life annuity business, assessing solvency has to be framed within a wide time horizon, where specific financial and demographic risks are realized. In this order of ideas, solvency indicators have to

capture the amount of capital to cope with the impact of those risk sources over the considered period.
We present a study of the dynamics of the solvency ratio, measuring the portfolio surplus in relation to its variations on fixed time intervals; these variations are restyled according to a risk-adjusted procedure.
Keywords: Life annuity, Solvency Ratio, Risk-Adjusted Management

Multi-State Model for Evaluating Conversion Options in Life Insurance

Guglielmo D'Amico[1], Montserrat Guillen[2], Raimondo Manca[3], Filippo Petroni[4]

[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy, [2]Dept of Econometrics, Statistics and Economics, University of Barcelona, Spain, [3]Dept of Methods and Models for Economics, Territory and Finance, University "La Sapienza", Italy, [4]Dept of Business, University of Cagliari, Italy

The conversion option is an option that allows the policyholder to convert his original temporary insurance policy (TIP) to permanent insurance policy (PIP) before the initial policy is due. In this work we propose a multi-state model for the evaluation of the conversion option contract. The multi-state model is based on generalized semi-Markov chains that are able to reproduce many important aspects that influence the valuation of the option like the duration problem, the non-homogeneity and the age effect. Finally, a numerical example shows the possibility of implementing the model in real-life problems.
Keywords: multi-state model, actuarial evaluation, life insurance.

Study of Human Migration into EU Area: A Semi-Markov Approach

Guglielmo D'Amico[1], Jacques Janssen[2], Raimondo Manca[3], Donatella Strangio[3]

[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy, [2]Honorary professor Universitè Libre de Bruxelles, Belgium, [3]MEMOTEF University of Roma "La Sapienza", Italy

It is well known that the migration models could be well studied by means of a semi-Markov process because this tool permits to take into account of the flux but also of waiting time, in this case, inside a country.
In this period, given the serious political problems in African and Middle East countries, the migration into some countries of EU increased in a

substantial way. In this study, we will classify the countries that are interested to this phenomenon in starting, transient and arriving countries. We will also take in great relevance the mean waiting times in each transient and arriving state. We will give also the probabilities of migration among the interested countries and, furthermore, the calculation of the mean time that is necessary for the arrival into the final destination.

Detecting Change-Points in Indexed Markov Chains with Application in Finance

Guglielmo D'Amico[1], Ada Lika[2], Filippo Petroni[3]

[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,
[2]Dept of Business, University of Cagliari, Italy,
[3]Dept of Business, University of Cagliari, Italy

We study the high frequency price dynamics of traded stocks by a model of returns using an indexed Markov approach. More precisely, we assume that the intraday returns are described by a discrete time homogeneous Markov model which depends also on a memory index. The index is introduced to take into account periods of high and low volatility in the market. We consider the change of volatility as the changing point for the Indexed Markov chain. In this work we present a method to detect these changing points and we apply the method to real data. In particular, we analyzed high frequency data from the Italian stock market from first of January 2007 until end of December 2010.
Keywords: high-trequency finance, bootstrap, change-points.

Volatility Forecasting by Means of a GARCH Model: Accuracy, Entropy and Predictability

Guglielmo D'Amico[1], Filippo Petroni[2], Flavio Prattico[3]

[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,
[2]Dept of Business, University of Cagliari, Italy, [3]Dept of Methods and Models for Economics, Territory and Finance, University "La Sapienza", Italy

The conversion option is an option that allows the policyholder to convert his original temporary insurance policy (TIP) to permanent insurance policy (PIP) before the initial policy is due. In this work we propose a multi-state model for the evaluation of the conversion option contract. The multi-state model is based on generalized semi-Markov chains that are able to reproduce many important aspects that influence the

valuation of the option like the duration problem, the non-homogeneity and the age effect. Finally, a numerical example shows the possibility of implementing the model in real-life problems.
Keywords: multi-state model, actuarial evaluation, life insurance.

A Dynamic Approach to the Modeling of Poverty

Guglielmo D'Amico[1], Philippe Regnault[2]

[1]Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,
[2]Laboratory of Mathematics, University of Reims Champagne-Ardenne, France

In this paper we extend some of the classical poverty indexes into a dynamic framework using continuous time Markov systems. The dynamic indexes are then generalized to interval based indexes and they are evaluated both in the case of a finite population and of an infinite population. The estimation methodology is presented under different sampling schemes and a simulation based example illustrates the results.
Keywords: Poverty estimation, Markov systems, dynamic indexes.

New Distribution Families for Modeling the Impact Factor

Silvia Dedu[1], Vasile Preda[2]

[1]Dept of Applied Mathematics, Bucharest University of Economic Studies and School of Advanced Studies of the Romanian Academy, Bucharest, Romania
[2]Bucharest University of Bucharest, Faculty of Mathematics and Computer Science and Romanian Academy, National Institute for Economic Research, Bucharest, Romania

Performance evaluation of scientific research represents a topic of real concern in the recent literature. The study of informetric distributions, such as distributions of citations and impact factors is one of the most relevant issues in the current informetric research. Recently, several laws for modeling the impact factor, based on ranks, have been proposed, including Zipf, Lavalette and the two-exponent law proposed by Mansilla et al. in 2007. In this paper we propose a new family of distributions which extends recent results obtained in this field. The most relevant probabilistic descriptive measures for the new family are derived, including moments, quantiles, inequality measures and Gini index. Also, the distribution of the order statistics is derived and several estimation methods are performed, includng quantiles, least squares and

maximum likelihood estimation methods. The theoretical results and the effectiveness of our model are illustrated by using a database containing impact factors from some relevant scientific fields. The applications presented prove that the new distribution family allows a higher degree of flexibility for modeling the impact factor. Computational results are provided in order to highlight the advantages of our model.

Keywords: research evaluation, informetric distributions, impact factor modeling, inequality measures, estimation.

Acknowledgement: *This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2905.*

Risk Assessment and Survival Models Involving Truncated and Censored Random Variables Using Information Measures

Silvia Dedu[1], Vasile Preda[2], Muhammad Sheraz[3]

[1]Dept of Applied Mathematics, Bucharest University of Economic Studies and School of Advanced Studies of the Romanian Academy, Romania, [2]Bucharest University of Bucharest, Faculty of Mathematics and Computer Science and Romanian Academy, National Institute for Economic Research, Romania, [3]Dept of Mathematical Sciences & Department of Economics and Finance, Institute of Business Administration, Pakistan

Risk assessment and optimization represent important topics in various research fields. One of the main objectives which define an optimal strategy in actuarial science is related to risk minimization, where risk refers to the uncertainty of its claims. Entropy represents a fundamental concept used to evaluate the uncertainty degree associated with random variables or phenomena. It can be used as a measure of variability for continuous random variables or a measure of diversity
regarding the possible values of discrete random variables. Due to the widespread applicability, the derivation of closed expressions for various entropy measures corresponding to univariate and multivariate distributions has been a subject of interest. The aim of this paper is to develop an entropy based approach to risk assessment for actuarial models involving truncated and censored random variables. By using some general information measures, the effect of different partial insurance schemes upon the entropy of losses is investigated. Analytic expressions for the per-payment and per-loss entropies and relationships between these measures are obtained. Also, the entropies of losses corresponding to proportional hazards and proportional reversed hazards models are derived. The applications presented prove that information theory approach using general information measures for

loss models allows a higher degree of flexibility. Computational results are provided.

Keywords: risk assessment, information measures, truncation, censoring, loss models.

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

Time Series Analysis of Online Ratings Data

Yiannis Dimotikalis

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

In web sites like TripAdvisor or Google Play users/members encouraged to give their 5-star rating for a hotel, attraction, restaurant, android app, etc. Those categorical scalar ratings range from few decades to several millions for a particular “object”. The time evolution of those ratings data is a categorical time series and can represented as an integer valued time series. In this work certain time series of hotel ratings from around the world are analyzed by the techniques of Integer time series models approach. Because we strongly believe the Binomial distribution of those data frequencies we compare our results to simulated time series generated from the appropriate Binomial Distribution B(n,p). As fitting criterion, the false rate% used and tested. The main result is the oscillating behavior of the observed and simulated integer time series of ratings, some suggestions and outcomes also discussed.

Keywords: Time Series Analysis, Integer Time Series model, Binomial Distribution, False Rate, Online Rating, Non-Linear Regression, Five Stars Rating, TripAdvisor.

Numerical Results of Critical Stock Price for American Put Options with Exercise Restrictions

Domingos Djinja

Dept of Mathematics and Informatics, Faculty of Sciences, Eduardo Mondlane University, Mozambique

American options are commonly traded every world. It is known that there is no a closed formula to price an American put option. An implicit formula to price American put options with exercise restrictions on weekends was derived by Djinja (2015). However, the optimal exercise

boundary was found numerically by finite difference method. In this paper, we evaluate the critical stock price (the optimal exercise boundary) by solving numerically the corresponding implicit integral equation.

Consumer Durables Possession Information for the Household Situation Modelling

Marta Dziechciarz-Duda, Anna Król

Dept of Econometrics, Wroclaw University of Economics, Poland

The description of household situation may concentrate either on poverty (lowest income decile(s), quintile or tertile); average situation (median, medium quintile or tertile) or wealth concentration (concentration indices, highest income decile(s), quintile or tertile). The identifying of the household situation (wellbeing) usually takes into consideration its multidimensionality. Practically it means, that the study tries to capture three aspects: income and expenditure i.e. monetary measures, subjective income evalua-tions and dwelling conditions. Unfortunately, income-based measures of well-being do not capture differences over time or across households in wealth accumulation, ownership of durable goods or access to credit. Interesting approach of descriptive analysis of households' situation is material wellbeing measurement, where the information concerning durables possession is used. Measures of durable ownership and durable replacement expenditure strongly correlate with self-perceived measures of both social status and quality of life, which suggests an important role for household situation description. The difficulty here is that of interest is not just ownership but also the quality and age of durables, as this will affect the consumption benefits available from the good. A durable good is a consumption good that can deliver useful services to a consumer through re-peated use over an extended period of time. According to the System of National Accounts the distinction is based on whether the goods can be used once only for purposes of production or consumption or whether they can be used repeatedly, or continuously. Econometric techniques are promising tool for household situation modelling. Commonly used are multivariate regression analysis, probit (or logit), discriminant analysis and canonical analysis. In the paper, the results of an attempt to analyse factors of endowment with selected consumer durables in Poland will be described.

Keywords: Durable goods, Households well-being, Multidimensional statistical methods.

Quartile Method Estimation of Two Parameter Exponential Distribution Data with Outliers

Entisar A. Elgmati[1], Nadia B. Gregni[2]

[1]*Dept of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya*
[2]*Dept of Statistics, Faculty of Science, Tripoli University, Tripoli, Libya*

Several methods have been used to estimate the unknown parameters in the two parameter exponential distribution. Here we have considered two of this methods, maximum likelihood estimates and median-first order statistics method. However, in the presence of outliers these methods are not valid. In this paper we propose two approaches that deal with this situation. The idea is based on using first and third quartile instead of the minimum statistics. We investigated the parameters estimate using these methods through simulation study. The new method gives similar results under the normal situation and much better results when the data has outliers.
Keywords: Exponential distribution, Maximum likelihood estimation, Median method, Quartile method.

Comparison of Complex and Simple Discriminant Analysis Approaches on Longitudinal Designs Data

Riham El Saeiti, Gabriela Czanner, Marta García-Fiñana

Biostatistics Dept, University of Liverpool, United Kingdom

Discriminant function analysis is often used to classify individuals into two or more groups. We propose a complex discriminant analysis approach when both longitudinal information (measurements taken over time) and covariate information (such as age, gender, etc.) are involved in the same model. One of the challenges is to construct appropriate covariance matrices that accounts for the correlations between measurements over time and cross-sectional covariates.
The complex method is done in two steps i) characterize the changes of the longitudinal markers via a multivariate linear mixed-effect model, then ii) use the multivariate model to derive linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) to predict the failure of treatment. On other hand, the simple method is to apply classical discriminant analysis approach (linear and quadratic) which require complete data to predict treatment failure.
Our approach will be applied to predict treatment success or treatment failure in patients with neovascular age-related macular degeneration at Paul's Eye Unit, Royal Liverpool University Hospital. The comparison

can be summarised into two main points: Compare between simple method that applied on balanced, completed design data (that approximated the time point) and complex method that applied on unbalanced, completed design data (that using the exactly time point). In Addition, examine the effect approximating the true time of patients follow in the clinic.
The second comparison is compare between simple method that applied on balanced, completed, imputation design data and complex method that applied on unbalanced, uncompleted design data (using the exactly time point).
Approximating the time points to have a balanced and completed design dataset does not seem to provide a less or more accurate prediction. The classification using complex method increases the AUC to approximately 94% compare to simple method.
Keywords: Discriminant function analysis, multivariate linear mixed-effects model

Hidden Markov Change Point Estimation

Robert J. Elliott[1], Sebastian Elliott[2]
[1]University of Adelaide and University of Calgary, [2]Elliott Stochastics Inc, Australia

A hidden Markov model is considered where the dynamics of the hidden process change at a random `change point' tau. In principle this gives rise to a non-linear filter but closed form recursive estimates are obtained for the conditional distribution of the hidden process and of tau.
Keywords: Hidden Markov Model, filter, recursive estimates

Using Graph Partitioning to Calculate Pagerank in a Changing Network

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

PageRank is a method which is used to rank the nodes of a network such as the network consisting of the webpages on the Internet and the links between them. Many real world networks change over time resulting in the need of fast methods to re-calculate PageRank after some time.

In this talk we will show how the old rank and a partition of the network into strongly connected components can be used to find the new rank after doing some certain types of changes in the network. In particular, three types of changes will be considered: 1) changes to the personalization vector used in PageRank, 2) adding or removing edges between strongly connected comnponents and 3) merging of strongly connected components.
To achieve this a partition of the network into strongly connected components together with a non-normalised variation of PageRank based on a sum of random walks on the network will be used.
Keywords: PageRank, graph, random walk, strongly connected component.

Clustering Method for Collaborative Filtering Applied to Applicant's Preference Determination in the Informational System E-Admiterea of the Moldova State University

Natalia Eni, Igor Bogus, Florentin Paladi

Moldova State University, Moldova

Collaborative filtering - is a method that gives automatic projections based on existing information on the interests and tastes of the users. We have implemented an approach which provides guidance on the selection of applicants matching their specialties, using the parameters of a statistical model to estimate the preferences. To elaborate a statistical model, we used the method of cluster analysis.
Informational system e-Admiterea was designed to automate business processes for enrolment of students to the Moldova State University. Selecting specialties by applicants is an important element in their future professionalization. One of the system's functions is to support the applicant in the selecting specialties taking into account his options and online submission of documents. Therefore, data on applicants are introduced online by the applicants themselves.
The paper propose is to analyze the data stored in the information system e-Admiterea for decision during the specialty selection based on statistics from previous two years, and statistical model building based on clustering analysis method. The preferences of each applicant are shown by a vector in 75-dimensional space (the number of spaces equals to the number of specialties), where projection on any axis is equal to 1, if applicant selected corresponding profession and 0 - otherwise. Then, using the clustering analysis one finds weights to each applicant's neighbors and calculates by collaborative filtering recommendation when choosing the suitable specialties for each candidate.

Keywords: collaborative filtering, clustering, e-Admiterea

Testing for Co-bubble Behaviour in Economic and Financial Time Series

Andria C. Evripidou

School of Economics, University of Nottingham, United Kingdom

The efficacy of unit root tests for detecting explosive rational asset price bubbles is well documented. The possibility of co-bubbling behaviour of two series is, however, less understood. We develop a methodology for testing the hypothesis of co-bubbling behaviour in two series, employing a variant of the stationarity test of Kwiatkowski et al. (1992) which uses a conditional `wild' bootsrap scheme to control size. Monte Carlo simulations offer promising levels of size control and power. Combining this test with a recently proposed Bayesian Information Criterion model selection procedure to identify bubble episodes in individual series allows us to determine the presence of any relationship between two explosive series robustly. An empirical application involving world silver and gold prices is presented.

Truncated Negative Exponential Distribution

Farmakis Nikolaos, Papatsouma Ioanna

Dept of Mathematics, Aristotle University of Thessaloniki, Greece

The basic version of the negative exponential distribution with parameter λ is a very useful and very often used distribution, connected with a great deal of Socio-economic, Political, Medical or Biological issues. A random variable X (rv X) with the above distribution takes its values in $\boldsymbol{R}_+$. In this paper we deal with rv X *having some kind of* truncated version of the above exponential distribution, defined in the space $[0, \beta]$. Several parameters of that distribution are studied and an effort takes place in order to define an estimator of the probability density function (pdf) via sampling procedures. Starting from the range β and some assumptions, all the other parameters are estimated, i.e. the basic parameter λ of the distribution, and a kind of its inflation rate $c>0$, etc. As the inflation rate is adopted, the range becomes finite and so the distribution becomes truncated. The Coefficient of Variation (Cv) is also used for the suitable polynomial approximation of pdf (Cv-methods). The exponent of the polynomial distribution is calculated directly from the Cv. This last version is used in order to compare (and connect) the exponential and

the polynomial character of the estimators of the distribution. The polynomial version of the distribution is more flexible and easier to be used for any study on the distributions of random variables.

Keywords: Exponential Distribution, Truncated distribution, Sampling, Random Variable.

AMC2010 Classification*:* 62D05, 62E17

REFERENCES

1. **Cochran, W.,** (1977). *Sampling Techniques*, 3rd edition, John Wiley & Sons, New York.
2. **Farmakis, N.,** (2001). *Statistics: Short Theory-Exercises*, 2nd edition, A&P Christodoulidi Publishing Co, Thessaloniki. (In Greek)
3. **Farmakis, N.,** (2003). Estimation of Coefficient of Variation: Scaling of Symmetric Continuous Distributions, *Statistics in Transition*, Vol. **6**, No 1, 83-96.
4. **Farmakis, N.,** (2006). A Unique Expression for the Size of Samples in Several Sampling Procedures, *Statistics in Transition*, Vol. **7**, No. 5, pp. 1031-1043.
5. **Farmakis, N.,** (2009a). *Introduction to Sampling*, A&P Christodoulidi Publishing Co, Thessaloniki. (In Greek)
6. **Farmakis, N.,** (2009b). *Surveys & Ethics*, 2nd edition, 2nd edition, A&P Christodoulidi Publishing Co, Thessaloniki. (In Greek)
7. **Kolyva-Machera, F., Bora-Senta, E.,** (2013). *Statistics: Theory and Applications*, 2nd edition, Ziti Publishing Co, Thessaloniki. (In Greek)

A Population Evolution Model and its Applications to Random Networks

István Fazekas, Csaba Noszály, Attila Perecsényi

University of Debrecen, Faculty of Informatics, Hungary

To describe real-life networks, the preferential attachment model was introduced by Barabási and Albert [1]. Then it was proved 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 (asymptotic) degree distribution. Following the paper of Barabási and Albert [1] several versions of the preferential attachment model were proposed. In Ostroumova *et al.* [4] a general graph evolution scheme was presented which covers lot of preferential attachment models. It was proved that the general model leads to a scale-free graph.

In this paper we present a further generalization of the model by Ostroumova *et al*. [4]. We consider the evolution of a population where the individuals are characterized by a score. During the evolution both the size of the population and the scores of the individuals are increased. We prove that the score distribution is scale-free. Then we apply our results to a random graph which is based on N-interactions (for the N-interactions model see Fazekas and Porvázsnyik [3] or Fazekas *et al*. [2]. We obtain that in the N-interactions model the weight distribution of the cliques is a power law.

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. L. Ostroumova, A. Ryabchenko, E. Samosvat. Generalized preferential attachment : tunable power-law degree distribution and clustering coefficient. Algorithms and models for the web graph, 185–202, Lecture Notes in Comput. Sci., 8305, Springer, Cham, 2013.

Full Interaction Partition Estimation in Stochastic Processes

Fernández M., Garcia Jesus E., González-López V.A., Viola M.L.L.
University of Campinas, Brazil

Consider Xt as being a multivariate Markov process on a finite alphabet A. The marginal processes of Xt interact depending on the past states of Xt. We introduce in this paper a consistent strategy to find the groups of independent marginal processes, conditioned to parts of the state space, in which the strings in the same part, of the state space, share the same transition probability to the next symbol on the alphabet A. The groups of conditionally independent marginal processes will be the interaction structure of Xt. The theoretical results introduced in this paper ensure through the Bayesian Information Criterion, that for a sample size large enough the estimation strategy allow to recover the true conditional interaction structure of Xt. Moreover, by construction, the strategy is also capable to catch mutual independence between the marginal processes of Xt. We use this methodology to identify independent groups of series from a total of 4 series with a high financial impact in the Brazilian stock market.
Keywords: Multivariate Markov chains, Independence, Partition Markov models, Financial market, Bayesian Information Criterion.

A Study on Intelligent Dashboard Support or Decision Making In Business Courier

R. P. Ferreira, A. Martiniano, A. Ferreira, K. R. Prado, R. J. Sassi
Nove de Julho University, Brazil

The aim of this paper was to research, evaluate and present a study on an intelligent dashboard to support decision making in courier company based on Artificial Intelligence techniques. Brazil has gone through

several transformations of general services have been adapting to the new demands of customers and market. As a result, the courier service has become highly complex and competitive. Transport, treatment and distribution remained follow these trends. In this context, the application of intelligent techniques to support decision-making is an alternative, seeking productivity and high level of service. The methodological synthesis of the article is to develop a dashboard supported by artificial intelligence techniques. An Artificial Neural Network (ANN) type Multilayer Perceptron (MLP), trained by error back-propagation algorithm was developed and applied to perform demand forecasting and prediction of absenteeism, these forecasts were presented in intelligent dashboard to support the making decision. Additionally we applied the Self-Organizing Map of Kohonen to generate groups seeking better visualization to be used on the dashboard. The data for the experiments were collected in a courier company. It was concluded that the application of techniques helped in the creation of an intelligent dashboard to support decision making.
Keywords: Dashboard intelligence, decision making, artificial neural networks, courier company.

A Study on Magic Square Applying Artificial Neural Networks

R. P. Ferreira, A. Martiniano, A. Ferreira, Al. Ferreira, R. J. Sassi

Nove de Julho University, Brazil

Magic Squares are formed by consecutive natural numbers, so that in all rows, columns and diagonals main summed up result in the same number, called the magic constant, the number of houses in a row is the square of the order. Artificial Neural Networks (ANN) models are made of simple processing units, called artificial neurons, which calculate mathematical functions. These models are inspired by the structure of the brain and aim to simulate human behavior, such as learning, association, generalization and abstraction when subjected to training. The aim of this paper is to apply a ANN to recognize the magic constant and the core value of the magic square. The ANN are particularly effective for mapping input / output nonlinear systems and to perform parallel processing, and simulate complex systems. ANN in the training phase hit 76% of the magic constants and 74.7% of the core values. In the test phase ANN hit 80% of the magic constants and 80% of the core values of the magic square. The Artificial Neural Network could recognize 80% of the results in the testing phase, which initially indicates an option to be used in this type of problem. It follows, therefore, that the purpose of the Article has been reached. As a future study aims to expand the tests with the magic squares in order to verify that the ANN

can have similar results using magic squares of different orders in the same database. Envisions with the development of the research presented in the article, the possible use in the recovery and/ or digital images encryption and certification.
Keywords: Magic square, Artificial Neural Network. Pattern Recognition.

A Study using Data Mining with Density Algorithm Based Scan Algorithm with Noise (DBSCAN) in Clustering Profiles of Employees Absenteeists and Presenteeists

R. P. Ferreira, A. Martiniano, R. J. Sassi
Nove de Julho University, Brazil

The advance of computational science allows processing of large amounts of data, which motivates the study. The term data mining has arisen due to the similarities between the search for important information in a database and the act of mine the mountain to find a vein of gold. Data mining is the element responsible for the efficient extraction of implicit and useful knowledge contained in a database. The purpose of this article is to apply data mining techniques to identify by clustering the profile absenteeist employees and presenteeist using DBSCAN algorithm. The Absenteeism is not reporting to work as scheduled. Presenteeism in there the employee's presence at work, even if sick, however, its activities are unproductive. The DBSCAN algorithm is applied in data mining to cluster. The results presented by DBSCAN algorithm showed that the application in data mining is satisfactory, confirming the use of this technique as an option to apply this type of problem. The methodology used in the structuring of this article was defined as literature, exploratory and experimental.
Keywords: Absenteeism, Presenteeism, Data Mining, DBSCAN.

Applying Ant Colony Optimization and Kohonen Self-Organizing Map in Vehicle Routing and Scheduling Problems

R. P. Ferreira, A. Martiniano, A. Ferreira, Al. Ferreira, R. J. Sassi
Nove de Julho University, Brazil

A typical routing problem involves several stops, various types of vehicles and many scripts, which makes the search for solution a complex problem. Various techniques are applied to solve this problem, including artificial intelligence techniques. The objective of this study was to apply the ant colony optimization algorithm (ACO) and the self-

organizing map of Kohonen the routing problem and vehicle scheduling. Applied the ACO from groups (clusters) generated by the self-organizing map (SOM) in the routing path to be travelled by vehicles. The combined technical and program managed routing vehicles, pointing as another option to apply this kind of problem.
Keywords: Orders, Routing and Vehicle Scheduling, Optimization algorithm Ant Colony, Self-Organizing Map.

Container Crane Bridge Driver with a Neuro Fuzzy Network Operation

R. P. Ferreira, A. Martiniano, A. Ferreira, E. A. Pimentel, R. J. Sassi
Nove de Julho University, Brazil

A container crane has the function of transporting containers from one point (the ship) to another point (train, truck or another ship). The difficulty of this task lies in the fact that the container is connected to the crane by cables, causing an opening angle while the container is being transported, interfering with the operation at high speed due to oscillation that occurs at the end point, may cause accidents. The Fuzzy Logic is a mathematical theory that aims to allow the modeling of approximate way of thinking, imitating the human ability to make decisions in uncertain and inaccurate environments. The concepts of fuzzy logic can be used for translating mathematically inaccurate information expressed by a set of linguistic variables. The Fuzzy Logic uses the idea that all things are possible membership degrees and attempts to model the sense of words, taken by common decision or judgment. The Artificial Neural Networks (ANN) models are made of simple processing units, called artificial neurons, which calculate mathematical functions. These models are inspired by the structure of the brain and aim to simulate human behavior. The aim of this paper is to present a crane container controller pre-project using a Neuro Fuzzy Network (NFN), which is the combination of fuzzy logic and ANN.
Keywords: Container Crane bridge, Fuzzy Logic, Artificial Neural Networks.

Demand Forecasting with Artificial Neural Networks

R. P. Ferreira, A. Martiniano, A. Ferreira, Al. Ferreira, R. J. Sassi
Nove de Julho University, Brazil

In recent decades, Brazil has undergone several transformations, from a closed economy to a market economy. Transport, treatment and

distribution of orders remained follow these trends. As a result, the delivery parcel service has become highly complex and competitive. In this context, the forecast demand of orders comes as differential, leading structured productivity and high level of customer service. The paper aims to provide for the daily demand of orders in an Orders Treatment Centre for fifteen days using Artificial Neural Network (ANN). The methodological synthesis of the article is the development of a Artificial Neural Network Multilayer Perceptron (MLP), trained by error back-propagation algorithm. The data for the experiments were collected for 60 days, 45 days to training and 15 days for testing. Experiments were performed with ten different topologies of ANN by changing the following parameters: number of hidden layers, number of neurons in the hidden layers, learning rate, momentum rate and stopping criteria. The results obtained with use of ANN in daily demand forecast orders showed good adhesion to the experimental data in the training and testing phases.
Keywords: Demand Forecasting, Orders, Artificial Neural Network.

Some New Methods for Goodness-of-Fit

Andrey Feuerverger
Dept of Statistical Sciences, University of Toronto, Canada

In this poster session we present some new methods for goodness-of-fit (GoF) which are tailored to emphasize the behaviour in the tails of the distribution. We extend the methods first proposed by Anderson and Darling (1952, 1954), and later extended by Sinclair, Spurr and Ahmad (1990), and Chernobai et al (2007, 2012) to place greater emphasis in the tails. We examine the behaviour of these new GoF tests in some detail and propose new and computationally efficient procedures based on saddlepoint approximations and jackknife methods for determining p-values when testing for membership in a parameterized distributional family. A key motivation for this work is the need for more accurate procedures in areas such as assessing operational risk using operational loss data in banking, and similar commercial contexts. It is argued that such applications must be regarded as being primarily of a statistical nature and that they lie at the frontier of statistical research.
Keywords: Anderson-Darling upper tail tests; goodness-of-fit; saddlepoint approximations.

Differences in Mortality According to Marital Status in the Czech Republic

Tomas Fiala[1], Jana Langhamrová[2], Jitka Langhamrová [1], Jindrich Klufa[3]

[1]Dept of Demography, Faculty of Informatics and Statistics, University of Economics, Czech Republic, [2]Dept of Statistics, Faculty of Informatics and Statistics, University of Economics, Czech Republic, [3]Dept of Mathematics, Faculty of Informatics and Statistics, University of Economics, Czech Republic

The political and social changes in the Czech Republic started in 1989 were followed by changes in demographic behaviour, too. The life expectancy (the value of which in 1990 was for males almost the same like in 1960) started to rise again relatively rapidly and is growing until now.

The level of mortality depends not only on age and sex, but also on marital status. Analysis of this dependence is not so frequent because of problems of obtaining appropriate detailed data.

The paper brings basic analysis of mortality rates in the Czech Republic according to age, sex and marital status in the period 1990–2014. The development of mortality rates, life expectancy at births and at 65 years of age distinguished not only by sex but also by marital status are presented. The differences in mortality between males and females and between particular types of marital status are analysed.

The analysis is based on data provided by the Czech Statistical Office and Eurostat.

Keywords: Mortality, sex, age, marital status, education level, Czech Republic.

Escalating Project Costs and the Gumbel Distribution

Jim Freeman, Nuno Gil

Alliance Manchester Business School, University of Manchester, UK

Cost overruns are a pervasive phenomenon for many construction and engineering projects. Common experience is that the larger the project, the larger the potential overrun. In the case of London Crossrail – one of the most ambitious infrastructure projects ever undertaken in the UK – overruns on just three stages of the project have been found to range from 70% – 358%. Intervening cost hikes, correspondingly, can be shown to conform to a Gumbel distribution. The latter discovery enables realistic overrun probabilities to be computed – vital to effective contingency planning.

Keywords: Contingency, Crossrail, Gumbel distribution, Escalating costs, overruns

Consistent Proximity between N-Grams

Garcia Jesus E., González-López V.A.
IMECC-UNICAMP, Brazil

In this paper we compare two criteria of proximity between stochastic processes, the symmetrized relative entropy and a new one based on the Bayesian Information Criterion (BIC). To formalize both criteria, we use the optimal estimation of N-grams, which is given by the Partition Markov Models (PMM). According to PMM the state space of a N-gram, composed by sequences of size N-1, is split into parts and the elements in the same part of the partition activate the same random mechanism to select the next element in the sequence. The consistent estimation of the partition can be obtained by use of the BIC. Moreover, the criterion of proximity based on the BIC is consistent to decide if two N-grams are near or not. We used both criteria to investigate for differences and similarities between five DNA sequences corresponding to four complete, registered Epstein Barr Virus (EBV) sequences (B95-8-type I, GD1-type I, GD2-type 1, and AG876-type II) and a new EBV strain type I sequence reported in 2012, HKNPC1.
Keywords: N-grams; Markov models; Bayesian Information Criterion

A Different Approach of Box-Jenkins Analysis Outlined by Romanian Economic Growth and Income Inequality

Carmen C. Gheorghe, Vasile Preda, Silvia Dedu
National Institute for Economic Research, Romanian Academy

In this paper we introduce a new approach of the Box-Jenkins Analysis. We test the model for Romanian Economic Growth and obtain that rate is ARIMA (2,1,1). Using this model, we forecast income inequality.
Keywords: Box-Jenkins analysis; ARIMA (2,1,1); forecasting; income inequality; economic growth

Maintenance Cost and Inspection Sample Size Analysis for a Multiunit Reliability Model Which Large Number of Units

Guldal Guleryuz[1], Ebru Yuksel[2]
[1]Dept of Industrial Engineering, Hacettepe University, Turkey, [2]Dept of Business Administration, TOBB University of Economics and Technology, Turkey

Maintenance and replacement models for multiunit deteriorating systems fall into some categories of maintenance policies. As for other systems, maintenance policies for systems of independently operating units where units are subject to random failures mainly aim to improve system availability and reduce system downtime, while reducing the maintenance cost.
In this study we consider a maintenance policy for a multiunit system with n devices subject to random failures. System is inspected at fixed times by taking a random sample of same size from units that are assumed to be working. The time between each consecutive inspection is also constant. Inspections are not perfect. During inspections, each failed unit in the sample can be detected only by a probability β. If a failed unit detected in the sample, it is replaced immediately, if not detected it stays failed. For the system to be in working conditions, the number of failed units should not exceed a threshold value.
The main goal is to minimize the total cost occurring due to failed devices while keeping the system working. It is also preferred to minimize the sample size for each inspection and make less frequent inspections as possible. For large number of devices under heavy traffic conditions, asymptotic behavior of the system is considered and parameters are calculated to minimize total cost for given policy.
Keywords: Reliability, Maintenance, Multiunit Systems.

Extension of Simple Logistic Regression as Variable Selection Technique in SNP Data Analysis

Nazatulshima Hassan[1,2], Gabriela Czanner[1], Andrea L Jorgensen[1], Marta García-Fiñana[1]

[1]Dept of Biostatistics, University of Liverpool, United Kingdom,
[2]Malaysian Institute of Chemical and Bioengineering Technology, Universiti Kuala Lumpur (UniKL MICET), Malaysia

In SNP data analysis, a typical procedure is to use a variable selection procedure where the primary aim is to identify the most important SNPs to produce a reliable prediction. The variable selection problem is still not solved, which is due to complex correlation structure of SNPs that can hinder the power of identification of small to moderate genetic effects. Here, we propose a novel method called tSNR. Originally, Signal-to-Noise Ratio (SNR) is defined as the ratio of the squared amplitude or variance of a signal relative to the variance of the noise. We extend the concept of SNR by calculating the ratio of deviance from a Simple Logistic Regression as the new ranking measure for Single Nucleotide Polymorphism (SNP) selection. The selected SNPs are then used as the

predictors for a classifier. We measure the efficiency of the method by calculating the classification accuracy, sensitivity, specificity and area under ROC curve (AUC).
Keywords: Variable Selection, Simple Logistic Regression, Signal-to-Noise Ratio, Single Nucleotide Polymorphisms (SNP).

Analysis of Evolution of Scores of the Program for International Student Assessment Through Analysis of Information

Valérie Girardin, Justine Lequesne, Olivier Thévenon

Laboratoire de Mathématiques N. Oresme, UMR6139, Campus II, Université de Caen Normandie, France, Organisation de coopération et de développement économiques (OCDE), Paris;
Institut national d' études démographiques (INED), Paris, France

In the continuation of SMTDA 2014, the setting of this work is the adjustment of a reference probability to given margins in a multidimensional contingency table in terms of discrimination information. A new log-linear parametric model -- that we call Zighera's model -- is obtained by repeated decomposition of Kullback-Leibler divergence.

By construction, its parameters are meaningful in terms of both information and odds ratios, for testing simultaneously the crossed influence of all explanatory variables, whichever be the dimension of the tables under study.

The Program for International Student Assessment (PISA) of the OECD studies the influence of time and personal background on the observed performance inequalities between school students in many countries.

The present contribution details an application to PISA data from 2000 to 2012 in France and Germany for mathematics literacy and reading literacy. In particular, Zighera's model allows one to highlight the evolution in the composition of the observed population that may affect the observed evolution of performance of students.

Keywords: information, log-linear model, odds-ratios, PISA.

References

1. V. Girardin, J. Lequesne and O. Thévenon. Décomposer l'information pour expliquer? Nouvelle approche log-linéaire des relations entre variables qualitatives et application aux enquêtes PISA, *Technical Report LMNO UCN*, 2016.
2. V. Girardin, J. Lequesne and A. Ricordeau. Discrimination information for analysis of contingency tables by log-linear models. *Rapport de recherche, Université de Caen Basse Normandie*, 2015.
3. V. Girardin. Analysis of information into marginal effects. *Communication SMDTA 2014*, Lisbon, Portugal.

Data Driven Modelling of Thermal Systems in Electric and Hybrid Vehicles

Gregor Gregorčič, Gernot Herschold-Pliessnig and Peter Drage

qpunkt GmbH, Gewerbepark 11, A-8075 Hart bei Graz, Austria

An electrification in automotive industry brings a new challenge in thermal management of the vehicle. Thermal systems in electric and hybrid vehicles operate with a much higher number of components than systems in conventional vehicles. While thermal management of the combustion engine, transmission and the passenger cabin is implemented in well-known and proven way, the integration of electric motors, power electronics and a battery introduces new functions in thermal management. The battery for example needs to operate at its well defined temperature level. The electric motor on the other hand operates at different temperature level and the power electronics operates at yet another temperature level. The temperature and the humidity in the passenger cabin must assure a high level of comfort. This requires an air conditioning system, which again operates at different temperature level. Multiple levels of operational temperatures require an effective thermal management of the overall vehicle. In order to analyse and develop effective control paradigms, simulation tools need to be extensively employed. Simulation demands accurate dynamic models of thermal cycles, which are highly nonlinear. Due to the complexity of such processes, first principle models and their analytical approximations cannot be efficiently applied and hence a more flexible type of model needs to be utilised. This talk presents modelling paradigms used to efficiently model thermal processes in electric and hybrid vehicles. A merge of probabilistic Gaussian process model and parametrical local model networks will be presented. The model structure and the model parametrisation will be explained. Since the proposed architecture assumes a data driven model, the importance of design of experiments for data collection will be highlighted. The utilisation of proposed model for development of advanced control strategies will be shown. Benefits of model based optimisation for development cost reduction will also be discussed.

On Strategically Equivalent Contests

Jean-Daniel Guigou[1], Bruno Lovat[2], Jang Schiltz[1]
[1]*University of Luxembourg, LSF, Luxembourg,*
[2]*University of Lorraine, BETA, France*

A contest is a game in which players expend costly resources, such as effort, money or time, to win a prize. The prize can be indivisible or it may be divisible. In the former case, expending resources increases the probability to win the prize (probabilistic contest). In the latter case, expending resources increases the share of the prize (deterministic contest). The equivalence of the probabilistic and deterministic contests can be derived under the assumption or risk neutrality and/or riskless prize. But we also show that the two-type contests may be strategically equivalent (in the sense that they both generate the same equilibrium expenditures) under risk aversion and risky prize.
This problem gives rise to a nonlinear, nonlocal, inhomogeneous differential equation. We show that there are two families of solutions, the exponential distribution, for any positive parameter λ, but also a family of densities with affine Laplace transforms.
Keywords: Game theory, Differential equations, Laplace transform.

A Niche-Structured Individual Based Model with Connectivity

Marc Joannides
IMAG, Universite de Montpellier, France

We investigate a continuous time stochastic individual-based model for a niche-structured population. Competition among individuals occurs only within the niches but immigration from neighboring niches is allowed. We give the stochastic differential equation governing the dynamics of the process and provide a macroscopic deterministic approximation by moment closure. The microscopic model and its approximation are compared through simulations.
Keywords: individual based model, Markov process, moment closure, ecology.

Algorithms for Numerical Simulation of Piecewise-Constant Periodically Correlated Processes Based on Different Types of Point Flows

Nina A. Kargapolova[1,2]

[1]*Novosibirsk National Research State University, Russia,*
[2]*Institute of Computational Mathematics and Mathematical Geophysics, Russia*

In this paper three models of asymptotically periodically correlated piecewise-constant non-Gaussian processes are considered. First model is based on heterogeneous Poisson point flows with periodically time-dependent intensity. As a basis of the second model heterogeneous Markov chains with transition probability matrixes, that are periodic functions of time, are used. The last model is based on regular point flows and special stochastic filling technique. For all considered processes analytical formulas, describing their one-dimensional distributions, mathematical expectations and correlation function, are given.
Keywords: Heterogeneous Markov Chain, Poisson Point Flow, Non-Gaussian Process, Periodically Correlated Process.
Acknowledgment: *This work was supported by the Russian Foundation for Basis Research (grants No 15-01-01458-a, 16-31-00123-mol-a, 16-31-00038-mol-a).*

Multivariate Analog of some Univariate Discrete Distributions

Yury Khokhlov

Lomonosov Moscow State University, Dept of mathematical Statistics, Russia

It is well known the role of some discrete distributions such as Bernoulli, binomial, geometric, negative binomial, Poisson, Polya-Aeppli and others in applications of probability theory and mathematical statistics. We propose some variant of multivariate distribution whose components has a given univariate discrete distribution. In fact we consider some very general variant of so called reduction method. We find the explicit form of mass function and generating function of such distribution, investigate its properties. We prove that our construction is the unique in natural exponential families of distributions. Our results are the generalization and unification of many results of other authors. Also we consider some applications of these distributions in actuarial and financial mathematics and teletraffic theory.
Keywords: Multivariate discrete distributions

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

The Population Projections under Different Sex Ratio at Birth Scenarios in Azerbaijan

Ismet Koc, Mehmet Ali Eryurt

Hacettepe University, Institute of Population Studies, Turkey

This paper aims to understand the impact of different levels of sex ratio at birth on age and sex structure of future population in Azerbaijan. In the study, the impact of skewed sex ratio at birth on size and structure of Azerbaijani population and marriage market were examined using the cohort component population projection methodology. Population projections, which were realized from 2014 to 2050 by using a fixed fertility, mortality and international migration schedules, based on five different sex ratios at birth scenarios. Five different scenarios of the sex ratio at birth (SRB) are: (1) Constant SRB scenario at the level of 116; (2) Slightly decreasing SRB scenario from 116 to 110; (3) Rapidly decreasing SRB scenario from 116 to 105; (4) Slightly increasing SRB scenario from 116 to 120; (5) SRB at the biological normal level scenario from 105 to 105.
Keywords: Son preference, population projection, Azerbaijan

Modeling of Parallel Development of Bond Market Indices

Jozef Komorník[1], Magda Komorníková[2], Tomáš Bacigál[2], Cuong Nguyen[3]

[1]Faculty of Management, Comenius University, Slovakia,
[2]Faculty of Civil Engineering, Slovak University of Technology, Slovakia,
[3]Faculty of Commerce, Lincoln University NZ, New Zealand

Stock and bond markets co-movements have been studied by many researchers. The object of our investigation is the development of three U.S. investment grade corporate bond indices. We concluded that the optimal 3D as well as partial pairwise 2D models are in the Student class with 2 degrees of freedom (and thus very heavy tails) and exhibit very high values of tail dependence coefficients. Hence the considered bond indices do not represent suitable components of a well diversified investment portfolio. On the other hand, they could make good candidates for underlying assets of derivative instruments.
Keywords: Bond market indices, Copulas, Cramer-von Mises test statistics, GoF test, Vuong and Clarke tests.

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

Asymptotic Ruin Probabilities for a Multidimensional Renewal Risk Model with Multivariate Regularly Varying Claims

Konstantinides Dimitrios, Li Jinzhu

University of the Aegean, Greece

This paper studies a continuous-time multidimensional risk model with constant force of interest and dependence structures among random factors involved. The model allows a general dependence among the claim-number processes from different insurance businesses. Moreover, we utilize the framework of multivariate regular variation to describe the dependence and heavy-tailed nature of the claim sizes. Some precise asymptotic expansions are derived for both finite-time and infinite-time ruin probabilities.

Keywords: asymptotics; multidimensional renewal risk model; multivariate regular variation; ruin probability

On the Use of Support Vector Machines for the Description of Demographic Patterns

Anastasia Kostaki[1], Javier M. Moguerza[2], Alberto Olivares[3], Stelios Psarakis[1]

[1]Dept of Statistics, Athens University of Economics and Business, Greece,
[2]Dept of Computer Science and Statistics, Rey Juan Carlos University, Spain,
[3]Dept of Signal Theory and Communications and Telematic Systems and Computing, Rey Juan Carlos University, Spain

The graduation of age-specific demographic rates is a subject of special interest in many disciplines as demography, biostatistics, actuarial practice and social planning. For estimating the unknown age-specific probabilities of the various demographic phenomena some graduation technique must be applied to the corresponding empirical rates, under the assumption that the true probabilities follow a smooth pattern through age. The classical way for graduating demographic rates is parametric modelling. However, for graduation purposes, nonparametric techniques can also be adapted. This work provides an adaptation, and an

evaluation of kernels and Support Vector Machines (SVM) in the context of graduation of demographic rates.
Keywords: graduation, mortality pattern, fertility pattern, kernels, Support Vector Machines

Simulation Based Reliability Evaluation of Maintenance Efficiency of Repairable System

Makram KRIT, Abdelwaheb REBAI
Tunisia

The aim of this paper is to study the asymptotic behavior of the Arithmetic Reduction of Intensity (ARI) and Arithmetic Reduction of Age (ARA) models as two imperfect maintenance models. These models have been proposed by Doyen and Gaudoin (2005), the failure process is simply Non Homogeneous Poisson Process (NHPP). The maintenance effect is characterized by the change induced on the failure intensity before and after failure during degradation period. To simplify study, the asymptotic properties of failure process are derived. Then, the asymptotic normality of several maintenance efficiency estimators can be proved in the case where the failure process without maintenance is known. Practically, the coverage rate of the asymptotic confidence intervals issued from those estimators is studied.

Quality Adjustments in Price Indexes for ICT Goods in Poland

Anna Król
Wroclaw University of Economics, Poland

The presented research addresses the problem of efficient methods of measuring actual price change of heterogeneous commodities which undergo very rapid technological development. On those markets it is very likely that certain difficulties arise in the process of prices comparison. Whenever there is significant change in characteristics of a commodity, new technology is implemented, or commodity is removed from the market, standard methods of price changes measurement might yield biased results, usually manifesting as overestimation of price growth rates. In response to this problem the hedonic approach, which links the price of the good ($PRICE$) with the set of its significant attributes (X) have been applied. The model, which represents above mentioned relationship ($PRICE = f(X; \alpha; \epsilon)$), allows to calculate theoretical price of a given good with specified set of characteristics, which in turn allows to

measure “true” price change. In presented research direct methods for calculating hedonic price indices have been applied for data from selected ICT market - tablet devices.
Keywords: Price indexes, Quality adjustments, Hedonic models, ICT market.

A Bayesian View on Some New Statistical Models for Early Life Failure Probability Assessment in Semiconductor Manufacturing

Daniel Kurz[1], Horst Lewitschnig[2], Jürgen Pilz[1]

[1]Dept of Statistics, Alpen-Adria-Universitaet Klagenfurt, Austria, [2]Infineon Technologies Austria AG, Austria

In semiconductor manufacturing, burn-in (BI) testing is applied to screen out early life failures before delivery. One approach to reduce the efforts of BI testing is to perform a BI study. That is, a random sample of devices is investigated for relevant failures after the BI. Based on that, an upper bound for the failure probability *p* of the devices in their early life can be assessed, typically by means of the Clopper-Pearson method. During the last years, we developed a series of new statistical models for assessing an upper bound for *p* on the basis of the binomial distribution. These models are capable of considering further available information, e.g.

- countermeasures implemented in the chip production process,
- the chip sizes of the individual products or
- BI studies on multiple reference products with different chip sizes.

Nevertheless, the main focus of these models is on the derivation of frequentist upper bounds for *p* on the basis of the Clopper-Pearson approach. In this talk, we discuss some of these models from a Bayesian point of view.

- At first, we show how to apply Bayesian inference for *p* under failures tackled by countermeasures. A beta Poisson-binomial distribution is applied to model the uncertainty in the effectivenesses of the countermeasures. Furthermore, we derive the Bayesian upper bound for *p* from a beta preposterior distribution.
- Second, we discuss how to perform area scaling for *p* from a Bayesian point of view. The posterior distribution for *p* of some reference product is assessed on the basis of the BI study. From this, the posterior distributions for the failure probabilities of follower products with different chip sizes can be inferred.
- Last but not least, we show how to combine BI studies on differently sized reference products from a Bayesian point of view.

We make use of the BI study of one reference product to construct a prior distribution for *p* of the second reference product. In this way, we can derive the posterior distribution for *p* of some follower product on the basis of both reference products.

In comparison to the frequentist approach, the main advantages of handling these new models from a Bayesian point of view are: less conservative upper bounds for *p*, consideration of possible expert knowledge about *p* and remedy to complex combinatorial derivations.

Keywords: Area scaling, Bayes, binomial distribution, burn-in, failure probability.

Acknowledgment: *The work has been performed in the project EPT300, co-funded by grants from Austria, Germany, Italy, The Netherlands and the ENIAC Joint Undertaking. This project is co-funded within the programme "Forschung, Innovation und Technologie für Informationstechnologie" by the Austrian Ministry for Transport, Innovation and Technology.*

References:

1. M.S. Hamada, A.G. Wilson, C.S. Reese, and H.F. Martz: Bayesian Reliability. Springer, New York, 2008.
2. D. Kurz, H. Lewitschnig, and J. Pilz: Decision-Theoretical Model for Failures Tackled by Countermeasures. IEEE Trans. Reliab., 63(2): 583-592, 2014.
3. D. Kurz, H. Lewitschnig, and J. Pilz: Failure Probability Estimation with Differently Sized Reference Products for Semiconductor Burn-in Studies. Appl. Stoch. Mod. Busin. Ind., 31(5): 732-744, 2015.

A Stochastic Vehicle Routing Problem with Ordered Customers and Stochastic Preferences and Demands

Epaminondas G. Kyriakidis, Theodosis D. Dimitrakos, Constantinos C. Karamatsoukis

Dept of Statistics, Athens University of Economics and Business, Greece

We develop and analyze a mathematical model for a specific vehicle routing problem that has many realistic applications. Specifically, we assume that a vehicle starts its route from a depot loaded with items of two similar but not identical products, which we name 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 number of items that each customer demands is a discrete 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 items 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 total expected 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: stochastic dynamic programming, vehicle routing problem

Threshold Models with Long Time Memory

Jean-Marie Le Goff

Centre for life courses and inequality studies (Lines), University of Lausanne, and Swiss National Center of Competencies in Research Overcoming Vulnerability: life course perspectives (Lives), Switzerland

Inverse Gaussian models are time to event models in which several underlined hypotheses are made (Whitmore 1983; Aalen & Gjessing, 2001; Lee & Whitmore 2010): first, for each individual, there is an underlined Brownian process. i.e. a continuous variable that follows a stochastic process that can be shaped by a normal law, with a linear drift; second, the event occurs when this underlined process reaches a threshold. This model is now used in a lot of quantitative disciplines in which time to event models are of interest, hypotheses of this model being interesting alternatives to the standard proportional hazard assumption (Lee & Whitmore 2010). In a certain manner, the inverse Gaussian model allow recognizing that the proportional assumption model is very arbitrary (Aalen et al 2008).
However, hypotheses of the inverse Gaussian model can be also discussed. Some authors criticize the arbitrary of the linear drift that do not reflect, for example, the process of death in human survival models for example (Jansen & Skiadas, 1995). In our communication, we are more interested to the aspect of Brownian motion. Since the 1950 years and work of Hurst (Hurst 1951; Hurst et al. 1965), it has been shown that a lot of stochastic processes have long-term memory, i.e., the value of the underlying process at one moment of the time is not completely random in respect the Gaussian nature of the process, but depends also from the past, even if its value remain random.

If Some authors have been interested to develop theoretically first passage models with long term memory (Rangarajan & Ding 2003), the approach proposed in our communication will is different. We are more interested to investigate behavior of the inverse Gaussian model when there is long term memory in the underlying process, or at the contrary when there is ante-persistence.

We simulate three samples of 2000 persons with underlying processes for the first sample following for each individual a standard Brownian process (Hurst constant H=0.5), in the second, a process with long term memory (H=0.8) and in the third, a ante persistent process (H=0.2). Simulations were built with the method of random midpoint displacement with random addition (Saupe, 1987). For each individual, the drift is considered to be null, while the variance of the random process, in absence of long-term memory or ante-persistence is fixed to one. For all individuals in each sample, the threshold is fixed to value 1, i.e, the event occurs when the level of the process becomes higher than one. Duration until the event is computed and a censor index of 1 is given. If the event is not reached after a duration of 85, this duration is taken into account with a censor index of 0.

Kaplan Meier estimations allow comparing the three samples. They show that events occur much more rapidly in the “ante-persistent” sample, while the process is much slower in the “long-term memory” sample. The speed of the “Brownian process” sample to experiment the event is intermediary. We then estimate on each sample two versions of the Inverse Gaussian model on each sample. In the first version, called the Whitmore version (Whitmore, 1983), is estimated the drift of the model and the standard error of the process, while the threshold is fixed to 1. In the second version, called the Gjessing version (Aalen & Gjessing, 2001), the standard error is fixed to 1 and the drift and the threshold are estimated. Models were estimated with the likelihood maximization technique.

Results show that whatever the version of the model and whatever the sample, the fit appear to be very good in a sense that fitting survival curve are very proximate to non-parametric Kaplan-Meier estimations. Moreover, the examination of the estimates of the drift show always an almost null drift. In the standard Brownian process model, the threshold is proximate to one in the Whitmore model, while the standard error is near 1 in the Gjessing model. In the case of the long term-memory model, the threshold is strongly overestimated (reaching a value 3.3) or the standard error is strongly underestimated. Results are opposite in the case of the ante-persistent model.

Such interesting results will be completed by simulations with non-null drifts, in order to better understand effect of long-term memory and ante-persistence on the behavior of the inverse Gaussian model.

Keywords: Threshold models, Long-term memory, Ante-persistence-Hurst constant, Simulations.

References :

1. Aalen O. O. & Gjessing H. K. (2001). Understanding the shape of the hazard rate. A process point of view. Statistical Science. 16(1):1-14.
2. Aalen A. O., Borgan H. & Gjessing H. (2008). Survival and event history analysis. A process point of view. New-York Berlin: Springer.
3. Hurst H.E. (1951). Long-term storage capacity of reservoirs. Transactions of the American society of civil Ingenieurs. 116: 770-808.
4. Hurst H.E, Black R.P. & Simaiki Y.M.. Long-term Storage. An experimental study. London: Constable.
5. Janssen J. and Skiadas C.H. (1995). Dynamic modelling of Life-Table data, Applied Stochastic Models and Data Analysis, 11, 1, 35:49.
6. Lee M.-L. T. & Whitmore G.A. (2010). Proportional hazards and threshold regression. Their theoretical and practical connections. Data Lifetime analysis. 16: 196-214.
7. Rangarajan G. & Ding W. (2003). First passage distributions for long memory processes. In Rangarajan G. & Ding W. (eds). Processes with Long-range correlations. Theory and applications 8pp 167-180). Berlin: Springer Verlag.
8. Saupe D. 81987). Algorithms for random fractal. In Peitgen H.O. & Saupe D. (Eds). The science of Fracatal Images (pp 71-136). New-York: Springer Verlag.
9. Whitmore G.A. (1983). A Regression Method for Censored Inverse Gaussian data. The Canadian Journal of statistics. 11(4): 305-315

New Directions for Goodness-of-Fit Methods and their Application

Victor Leiva, Claudia Castro-Kuriss

Adolfo Ibañez University, Viña del Mar, Chile

Goodness-of-fit methods are used for detecting whether a distribution is suitable to model a censored or uncensored data set. If the goodness of fit of a distribution to a censored data set is evaluated, then the methods must be adapted. In this talk we discuss some goodness-of-fit tests and their graphical counterparts. We conduct numerical studies for assessing the performance of these tests and for showing potential applications.
Keywords: Censored data; GOF tests; location and non-location-scale family; Michael test; probability plots

Bipartite Fuzzy Stochastic Differential Equations

Marek T. Malinowski

Institute of Mathematics, Cracow University of Technology, Poland

To handle dynamics of systems operating in random and vague/fuzzy environment, we propose to consider fuzzy stochastic differential equations. This constitutes a new branch of research in modeling uncertain phenomena. The papers [1—5, 9] contain studies on fuzzy and set-valued stochastic differential equations in a form that is a natural

extension of classical single-valued stochastic differential equations. In this case, solutions to fuzzy equations possess trajectories whose values become fuzzier and fuzzier as time increases. However, this kind of propagation of uncertainty is not desired always. Therefore, we studied the fuzzy stochastic differential equations of decreasing fuzziness in papers [6—8]. Now we combine two formulations building so-called "bipartite fuzzy stochastic differential equations". They should share the properties appropriate for both approaches with increasing and decreasing fuzziness. We examine existence and uniqueness of solutions to new equations. Also, stability properties of solutions are established. Several examples of particular bipartite fuzzy stochastic differential equations are presented. All the results apply immediately to bipartite set-valued stochastic differential equations.

Keywords: Fuzzy stochastic differential equation, modelling in fuzzy and random environment.

References

1. Malinowski, M. T., "Strong solutions to stochastic fuzzy differential equations of Ito type", Math. Comput. Modelling 55, 918-928 (2012).
2. Malinowski, M. T., "Ito type stochastic fuzzy differential equations with delay", Systems Control Lett. 61, 692-701 (2012).
3. Malinowski, M. T., "Some properties of strong solutions to stochastic fuzzy differential equations", Inform. Sciences 252, 62-80 (2013).
4. Malinowski, M. T., "Modeling with stochastic fuzzy differential equations", in: Mathematics of Uncertainty Modeling in the Analysis of Engineering and Science Problems, pp. 150-172, IGI Global, Hershey Pennsylvania (2014).
5. Malinowski, M. T., "Fuzzy and set-valued stochastic differential equations with local Lipschitz condition", IEEE Trans. Fuzzy Syst. 23, 1891-1898, (2015).
6. Malinowski, M. T., "Fuzzy and set-valued stochastic differential equations with solutions of decreasing fuzziness", Advan. Intellig. Syst. Comput. 315, 105-112 (2015).
7. Malinowski, M. T., "Stochastic fuzzy differential equations of a nonincreasing type", Commun. Nonlinear Sci. Numer. Simulat. doi:10.1016/ j.cnsns.2015.07.001, in press, (2015).
8. Malinowski, M. T., "Fuzzy stochastic differential equations of decreasing fuzziness: approximate solutions", J. Intellig. Fuzzy Syst. 29, 1087-1107, (2015).
9. Malinowski, M. T., Agarwal, R. P., "On solutions to set-valued and fuzzy stochastic differential equations", J. Franklin Inst. 352, 3014-3043, (2015).

Visualising Model Selection in Regression

Cecilio Mar-Molinero[1], Carlos Serrano-Cinca[2], Fabiola Portillo[3]

[1]University of Kent, UK, and University Autonoma of Barcelona, Spain,
[2]Dept of Economics and Business, University of Zaragoza, Spain,
[3]Dept of Economics and Business, University of La Rioja, Spain

Specifying a regression model requires deciding on which explanatory variables it should contain as well as deciding on the functional form that relates the dependent variable to the explanatory variables. Various model selection procedures exist. However, there still is no satisfactory

answer to the question of which model is "best". We propose to supplement model selection procedures with graphical representations obtained from the techniques of multivariate statistics. The figures obtained put in evidence the presence of extreme multivariate residuals, displays families of models with similar statistical performance, and can guide model selection for implementation purposes.
Keywords: Model averaging, Model selection, Multivariate analysis, Regression, Stepwise regression.

Evaluating Hog Prices in Mexico

Francisco Ernesto Martínez-Castañeda[1], Ana Lorga da Silva[2]

[1]Instituto de Ciencias Agropecuarias y Rurales, Universidad Autónoma del Estado de México, Mexico, [2]Departamento de Economia e Gestão, ECEO; CPES – ULHT, Portugal & CEDRIC-CNAM, France

In this work we study the evolution of Hog Price in Mexico since January 1987 until February 2015 in Mexico. We do univariate time series analysis, choosing the model that fits better this series, in order to forecast hog price series. The influence of Hog Price and it's production in Mexico's GDP it's also studied in the available period, but also the relation of Hog Price and Production -regression models are used. Hog prices are a common cyclical phenomena and it is related to different variables. Can be explained by technical factor, as well as economic variables even social variables. A basic principle in economics establishes that as offer of a product increase, prices diminish. Since 1938 with the Cobweb Theorem was described by Ezequiel, many constributions have been made, in order to describe and evaluate the influence of hog prices in production and viceversa.
Keywords: Regression Models, Time Series, Forecasting, Livestock Production.

An R Package for Discrete Beta Kernel Graduation of Probabilities of Dying

Angelo Mazza, Antonio Punzo

University of Catania, Dept of Economics and Business, Italy

Probabilities of dying are indicators commonly used in demography and actuarial practice. They are usually referred to one or more variables, the most common being age, calendar years and duration; for ease of presentation, we will start focusing on age only, while bivariate

extensions will be addressed afterward. To be specific, the dx deaths at age x can be seen as arising from a population, initially exposed to the risk of death, of size ex. This can be summarized via the model dx ∼ Bin (ex, qx), where qx represents the true, but unknown, mortality rate at age x. The crude rate °qx is the observed counterpart of qx. Graduation is necessary because crude data usually presents abrupt changes, which do not agree with the dependence structure supposedly characterizing the true rates (London, 1985). Nonparametric models are the natural choice if the aim is to reflect this belief. Furthermore, a nonparametric approach can be used to choose the simplest suitable parametric model, to provide a diagnostic check of a parametric model, or to simply explore the data (see H¨ardle 1990, Section 1.1, for a detailed discussion on the chief motivations that imply their use, and Debón et al. 2006 for an exhaustive comparison of nonparametric methods in the graduation of probabilities of dying).

Due to its conceptual simplicity and practical and theoretical properties, kernel smoothing is one of the most popular statistical methods for nonparametric graduation. Among the various alternatives existing in literature (see Copas and Haberman, 1983, Bloomfield and Haberman, 1987, Gavin et al., 1993, 1994, 1995 and Peristera and Kostaki, 2005), the attention is focused on the discrete beta kernel estimator proposed by Mazza and Punzo (2011). The genesis of this model starts with the consideration that, although age X is in principle a continuous variable, it is typically truncated in some way, such as age at last birthday, so that it takes values on the discrete set X = {0, 1, . . . , ω}, ω being the highest age of interest. Discretization could also come handy to actuaries who have to produce “discrete” graduated mortality tables. Discrete beta kernels are considered to overcome the problem of boundary bias, commonly arising from the use of symmetric kernels (Chen, 2000); their support X, in fact, matches the age range and this, when smoothing is made near the boundaries, allows avoiding allocation of weight outside the support (for example negative or unrealistically high ages). Adaptive variants, based on the reliability of the data at each age, are introduced in Mazza and Punzo (2013a,b).

In this poster we present the R (R Core Team, 2013) package DBKGrad (Mazza and Punzo, 2014c,b), available from CRAN (http://CRAN.R-project.org/), which implements all the methods described above and offers some related functionalities like bivariate graduation and diagnostic checks (cf. Mazza and Punzo, 2014a). Although R is well-provided with functions performing kernel smoothing techniques (see, e.g., Hayfield and Racine, 2008), there aren’t any offering discrete beta kernel smoothing or kernel smoothing in the field of graduation. However, there are packages that allow for nonparametric graduation using spline-based methods. For example, the demography package (Hyndman et al., 2012) does partially monotonic penalized spline smoothing for probabilities of dying and other demographic indicators.

Furthermore, the MortalitySmooth package of Camarda (2012) treats mortality data as Poisson counts and smooths using P-splines; similarly, to DBKGrad, this package also allows for bivariate graduation.

The Analysis of Student Paths at the University Using the Multivariate Joint Models

Marcella Mazzoleni

Dept of Statistics and Quantitative Methods,
Universita degli Studi di Milano Bicocca, Italy

The joint models analyse jointly survival and longitudinal data and they are composed of two sub-models, the longitudinal and the survival sub-model. It is interesting to analyse the situation in which one or both the sub-models are multi-variate. Firstly, it is possible to extend the longitudinal sub-model considering more than one longitudinal covariate. Thus a multivariate mixed model can be proposed for the longitudinal sub-model. Concerning the survival sub-model, a hazard model is proposed considering jointly the influence of all the longitudinal covariates onto the risk of the event. Secondly also the survival sub-model can be extended considering different competing or recurrent events. Thus in the survival sub-model different possible models are presented in order to deal with more than one event. Two possible solutions can be chosen, the use of different hazard functions for each competing terminal event, or the use of two kind of hazard function, one for the terminal and one for the recurrent events. Thus the aim of the joint models is to quantify the influence of one or more longitudinal covariates onto the risk of one or more events. The purpose of the paper is to find methods of estimation that are 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 application proposed analyses the undergraduates' paths in an Italian University, analysing the time of student graduation and the influence on the event graduation of one or more longitudinal covariate, such as the average grade and the number of exams that the student has already passed before a fixed time. Subsequently it is possible to analyse another competing event, the drop-out, which indicates the withdrawal of an undergraduate from the University system before the graduation. This event is considered as an opposite terminal event with respect to the graduation, then the hazard function is settle in a competing risk models framework.

Keywords: Joint Models, Multivariate Mixed Models, Time to Graduation, Competing Risk Models.

Bankruptcy Prediction Based on Bayesian Estimation of Logistic Regression Model with Random Effects

Sami Mestiri

Applied Economics and Simulation, Faculty of Management and Economic Sciences of Mahdia, Monastir University, Tunisia

The aim of this current paper is to predict the credit risk of banks in Tunisia, over the period (2000-2005). For this purpose, two methods for the estimation of logistic regression model with random effects: Penalized Quasi Likelihood (PQL) method and Gibbs Sampler algorithm are applied. By using information on a sample of 528 Tunisian firms and 26 financial ratios, we show that Bayesian approach improves the quality of model predictions in terms of good classification as well as by the ROC curve result.
Keywords: Forecasting, Credit risk, Penalized Quasi Likelihood, Gibbs Sampler, Logistic regression with random effects, Curve ROC.

Fertility Puzzle in India: Some Plausible Explanation

Sanjay K Mohanty

Dept of Fertility Studies, International Institute for Population Sciences, India

The secular decline in fertility trends in India along with improvement on distal determinants are encouraging. The TFR had declined from 3.4 in 1991 to 2.3 by 2011 and the declining trends is notice across socio-economic groups. The country has significantly improved in the key dimension of human development; education, health and income. However, the national trends in fertility conceal large disparities at state and districts of India. Recently released National Family Health Survey 4, 2015-16 revealed decline in use of contraception in 10 of the 16 states (for which preliminary results are available). It is puzzling to note the declining in fertility along with decline in both overall contraception and use of modern method in many of the states of India. Given that the induced abortion rate are low and the post partum insusceptibility has positive role in declining fertility, increasing age at marriage remains the only explanatory variables to explain such a puzzle. In this context this paper examines the distal and proximate determinants of fertility in India. Analyses is carried out both at state and district level for which data are available. Bogart's model is used to understand the role of proximate

determinants. Regression analyses has been carried out to understand the role of distal determinants of fertility decline in India.

PM10 Concentrations and Back Trajectories Analysis in the Amazon Rainforest after a Round Year

Daniel Moran-Zuloaga[1], Christopher Poelhker[1], Mira Poehlker[1], Jorge Saturno[1], Florian Ditas[1], David Walter[1], Isabella Hrabe de Angelis[1], Xuguang Chi[4], Paulo Artaxo[3], Meinrat O Andreae[2]

[1]Dept of Biogeochemistry and Multiphase, Aerosol Research Group, Max Planck Institute for Chemistry, Germany, [2]Scripps Institute of Oceanography, University of California San Diego, USA, [3]Instituto de Física, Universidade de Sao Paolo (USP), Brazil, [4]Nanjing University, China

The study of aerosol properties under natural conditions it is hard since there are few continental places around the world that suit these conditions. The Amazon rainforest in Brazil is well known as one of these natural hotspots. Amazon Tall Tower Observatory ATTO project is located in the middle of the Amazon rainforest, described by Andreae et al (2015). It is located in a very remote region with very few anthropogenic interventions occurring through the year.

Starting in 2014 until now a continuous measurement has been done to characterize the size distribution of aerosol particles from 0.06 to 10 µm, wide range from accumulation and coarse mode. To measure the size distribution, an Ultra High Resolution Aerosol Spectrometer (UHSAS) and an Optical Particle Sizer (OPS) with five minutes time resolution from 2014 and 2015. For the back trajectories, HYSPLIT trajectories every 14 days backwards and at 1000 m height were used. For statistical analysis: r project 3.2.0, rstudio 0.981103 and openair version 1.1-4, igor pro 6.0 were used.

Concentration Weight Trajectory CWT and Potential Source Contribution Function PSCF were used to resolve the back trajectories with the Particle Matter concentration through the time, the methodology is described by Carslaw and Beevers (2013)

The output showed defined seasonality behavior from accumulation and for coarse mode particles. Therefore, accumulation mode particles are easy to detect during the dry season [September to December]; while coarse mode particles are more detectable during the wet [February to April] and dry seasons. During the other months, Amazon rainforest experience a stage of cleanness, where very few external interventions are detected. Total number concentrations were > 800 #/cm3 for fine particles and and > 40#cm3 for coarse particles during the wet and dry season respectively. In contrast, for mass concentrations PM10 were higher during the wet season while PM1 increased during the dry

season; emphasizing the importance of the sources, external and local produced. There were also transitional periods (May-July) where low fine/coarse particles were found, in which less external pollution for local aerosol is release by the forest.

Brazilian rainforest is a very sensitivity place for external interventions usually classified as natural [wet season] and anthropogenic [dry season] produced. In a minor or greater degree, this seasonality effect is observed during the whole year. A cluster analysis was performed to better understand the emissions of the pollutants and to understand whether local or external occurred and when does it happened.

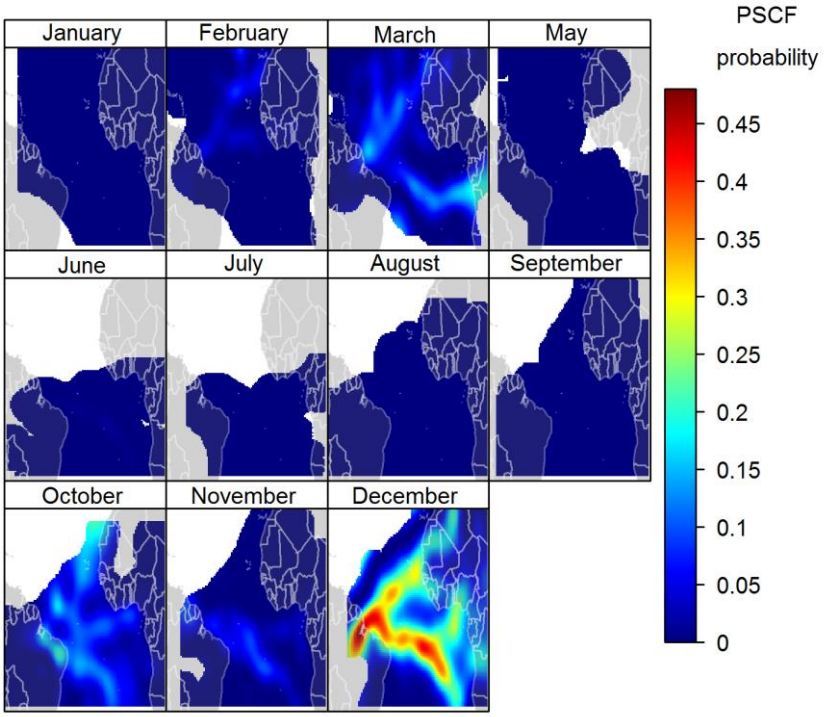

Figure 1. PM10 [colour bar units µg/m3] monthly concentrations calculated with Potencial Source Contribution Function PSCF (90th percentile) at ATTO site during 2015.

References:

1. Andreae, M. O, et al.: The Amazon Tall Tower Observatory (ATTO) in the remote Amazon Basin: overview of first results from ecosystem ecology, meteorology, trace gas, and aerosol measurements, *Atmos. Chem. Phys. Discuss*., 15, 11599-11726, 10.5194/acpd-15-11599-2015, 2015.
2. Carslaw, D. C., and Beevers, S. D.: Characterising and understanding emission sources using bivariate polar plots and k-means clustering, *Environmental Modelling & Software*, 40, 325-329, http://dx.doi.org/10.1016/j.envsoft.2012.09.005, 2013.

Impact of Early Pregnancy when Domestic Delivery is a Question for High MMR in India

Barun Kumar Mukhopadhyay
Indian Statistical Institute, India

A multivariate statistical analysis was recently conducted giving a significant contribution of domestic delivery (around 60 per cent) towards high maternal deaths (MMR) in India (Mukhopadhyay, 2015). From this a further objective naturally being made here as to whether early age at marriage followed by early pregnancy and delivery at home which is still persistent in India whether contributed to this high MMR. As per MDG of UN (2000) and later SDG (2012), the high MMR for India was to be reduced to 109 deaths/100,000 live births in 2015 of the former. The latest available figure from the SRS (GOI, 2015) is 167 in 2011-2013. A meeting among world leaders was held further by the UN for Sustainable Development Goal (SDG) especially for the developing countries. Still the estimated figure of MMR in India is only around 150 in the current year (Mukhopadhyay, op. cit). As a matter of fact, mere government macro level planning has not contributed effectively as India is a country of democracy and at the same time huge diversity among people living and freely moving to stay in different geographical region by their constitutional rights. My paper on the basis of data from different government and private sites tries to find the actual situation. The final result is yet to obtain although personal experience says that early marriages and infant and maternal deaths due to domestic delivery are evident even among elite class of population living in metro cities in India apart from huge masses living in remote villages.
Keywords: MMR, MDG, SDG, Age at marriage

Pricing European Options under Two-Dimensional Black-Scholes Partial Differential Equation by using the Crank-Nicholson Finite Difference Method

Jean-Paul Murara[1,3], Betuel Canhanga[2,3], Anatoliy Malyarenko[3], Ying Ni[3], Sergei Silvestrov[3]
[1]College of Science and Technology, School of Sciences, Dept of Applied Mathematics, University of Rwanda, Rwanda, [2]Faculty of Sciences, Dept of Mathematics and Computer Sciences, Eduardo Mondlane University, Mozambique, [3]Division of Applied Mathematics, School of Education, Culture and Communication, Malardalen University, Sweden

In the option pricing process, Black-Scholes in 1973 solved a partial differential equation and introduced a model to determine the price of European Options. Many researchers improved Black-Scholes model afterwards. Christoffersen proved in 2009 that models with two stochastic volatilities capture better the skewness and the smiles of the volatilities, meaning that they can more accurately determine the options prices. Canhanga et al. in 2014 and Chiarella and Ziveyi in 2013 used the model introduced by Christoffersen to determine European and American option prices respectively.
While dealing with many problems in financial engineering, the application of Partial Differential Equations (PDEs) is fundamental to explain the changes that occur in the evolved systems. Some families of this type of equations are known to have the so-called classical solutions. Others can be transformed into simple PDEs, for example by using scaling methods, Laplace and Fourier transforms, afterwards one can compute their solutions. Moreover, in many cases the PDEs that characterize the real life problems do not have known forms of solutions. In this occasion, numerical methods are considered in order to obtain the approximate solutions of the PDEs. In the present paper, we consider the option pricing problems that involves a two-dimensional Black-Scholes PDE as the one obtained by Canhanga et al. in 2014, and instead of solving it by the approximation approach presented by Conze in 2010 we perform the Crank - Nicholson finite difference method. Comparing examples are included in the paper.
Keywords: Stochastic Volatility, Two-dimensional Black-Scholes PDE, Crank-Nicholson Finite Difference Method.

Applications of Unbiased Post-Randomization to Control Identification Risk in Microdata Release

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

Statistical agencies aim to collect and publish useful data, but they also need to protect the confidentiality of unit level information for legal reasons and upholding public trust. So, agencies often release a perturbed or masked version of the original data. Developing suitable data masking methods is an important research topic.
In this talk, we shall first introduce a new measure of identification risk and a precise disclosure control goal. Then, we shall discuss several adaptations of unbiased post-randomization for achieving the stated goal. Specifically, we guarantee that the probability of correctly identifying the records of any target unit will be less than a prespecified

value, chosen by the agency. We shall also present some attractive properties of our approach. In particular, it is unbiased, in the sense that the expected frequency of any category under our procedure equals the original frequency. So, the usual estimators of the cell probabilities based on the original data remain unbiased when applied to perturbed data. Also, the variation of perturbed frequencies around the original values is quite small. Thus, the procedure does not have much negative impact on data quality.
Keywords: Confidentiality protection; identification risk; post-randomization; unbiased estimation; variance inflation.

Monte Carlo Algorithms for Markov Jump Processes and their Ergodicity Properties

Wojciech Niemiro

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Poland

Rao and Teh (2013) introduced an efficient MCMC algorithm for sampling from the posterior distribution of a hidden Markov jump process. The algorithm is based on the idea of sampling virtual jumps. In the present paper we show that the Markov chain generated by Rao and Teh's algorithm is geometrically ergodic. To this end we establish a geometric drift condition towards a small set. We also mention other algorithms of similar type.
Keywords: Continuous time Markov processes, MCMC, Hidden Markov models, Posterior sampling, Geometric ergodicity, Drift condition, Small set

Impact of a Conditional Cash Transfer Program on Infant Mortality Rates in the Brazilian Semiarid Region

Neir A. Paes[1], Everlane S. A. Silva[2]

[1]Dept of Statistics of the Federal University of Paraíba, Cidade Universitária, Brazil, [2]Federal University of Paraíba, Brazil

The Brazilian semiarid region is the world's largest in terms of density population and extension with 22 million inhabitants in 2010. Income inequality is among the highest in the world. In the past decade, Brazil has undergone notable social and expansion in public health, including a large reduction in infant mortality. In 2003, the federal government introduced the Bolsa Família Program (BFP) a widespread conditional cash transfer to poor households when they comply with conditions

related to health and education. A longitudinal ecological study using panel data from 1133 municipalities of the Brazilian semiarid region was performed. We aimed to assess the effect of the BFP on infant mortality in the semiarid region during the period 2004-2010. We employed conditional negative binomial regression models for panel data with fixed-effects specification at the municipal level. The primary outcome variable was the IMR, adjusted for relevant social and demographic covariates, and for the effect of the largest primary health-care scheme in the country through the Family Health Strategy (FHS). BFP and FHS had a significant role in reduction of infant mortality, reflected by the increase coverage of both programs. These government interventions added to increased coverage of prenatal care consultations and reduction of fertility levels had significant contribution to decreases in child mortality in the Brazilian semiarid region. The conditional cash transfers programs like BFP have a great deal of potential to improve population health, and target poor and vulnerable groups at greatest risk, particularly in the semiarid region, where can be found extremely poor people.

Keywords: Infant mortality, Cash Transfer Program, Brazil, Regression Analysis.

Which Effect Size Measure Should Be Reported for One-Way and Two Way ANOVA Models

Hüseyin Palazoğlu, Soner Yiğit, Mehmet Mendeş

Çanakkale Onsekiz Mart University, Biometry and Genetics Unit, Turkey

While reporting statistical analysis results, along with the P-value which shows statistical significance, reporting some effect size measures that show practical significance is quite important because effect size measures could show in both practical significance and how much variation in response could be explained by independent variables. In accordance with this purpose, recently, while reporting statistical analysis results, researchers have begun to report some effect size measures such as Eta-squared, Partial eta-squared, Omega-squared or Epsilon-squared along with P-value. The most important question at this point is determining that reporting which of effect size measure is more appropriate in terms of considered experimental conditions. In this study, effect size measures such as Eta-squared, Partial eta-squared, Omega-squared and Epsilon-squared are compared in terms of their performances. Simulatio! n results showed that the most unbiased estimations are obtained when Omega-squared and Epsilon-squared are used. It is seen that Eta-squared and Partial eta-squared, on the other hand, gives biased results in all experimental conditions. In conclusion,

although it is observed that Omega-squared and Epsilon-squared give slightly biased results in distributions with quite large kurtosis and skewness, this bias is negligible. Thus, while reporting results of both one-way and two-way analysis of variance, it is recommended that reporting Omega-squared and Epsilon-squared along with P-value is more appropriate.
Keywords: Effect size measure, simulation, biased, Omega-Squared, Epsilon-Squared

Pairwise Likelihood Estimation Based on a Sample of Pairs

Ioulia Papageorgiou, Irini Moustaki
Athens University of Economics and Business, Greece

Abstract Pairwise likelihood estimation has been recently developed for estimating the parameters of latent variable and structural equation models. Pairwise likelihood is a special case of composite likelihood methods that use lower order conditional or marginal log likelihoods. The composite likelihood to be maximised is a weighted sum of marginal or conditional loglikelihoods. Weights can be chosen to be equal or unequal for increasing efficiency. In this paper, we approach the problem of weights from a sampling perspective. More specifically, we propose a sampling method for selecting pairs that is based on those pairs that contribute more to the total variance from all pairs. We demonstrate the performance of our methodology using simulated examples.
Keywords: principal component analysis, structural equation models, confirmatory factor analysis.

Ignoring Non-Diagonalizable Matrices in Social Sciences

Pieter-Jan Pauwelyn, Marie-Anne Guerry
Dept of Business Technology and Operations (BUTO),
Vrije Universiteit Brussel, Belgium

In applied sciences, many transition processes are described by Markov models [1]. Markov chains use mainly stochastic matrices. The transition matrices consist of transition probabilities, coming from collected data. Thus, these transition probabilities are estimated values. In this paper, the interest lies with the non-diagonalizable stochastic matrices. The goal of this paper is to give an argument why we might neglect the non-diagonalizable stochastic matrices in social sciences.

This argument is based on perturbation theory and the fact that the diagonalizable stochastic matrices are dense in the set of the stochastic matrices [2]. We discuss the perturbation of stochastic matrices explicitly, the influence of every perturbation on a matrixelement will be evaluated. The evaluation will happen on eigenvalues, eigen-vectors, eigenspaces and systems of difference equations corresponding to a stochastic matrix [3]. This paper examines in specific 3x3 matrices. We won't consider the general case, but we give some possibilities for generalizing the result for 3x3-matrix.

Keywords: Markov models, Stochastic matrices, Non-diagonalizable matrices, Perturbation theory.

References

1. D.J. Bartholomew, Stochastic Models for Social Processes, 2nd edn, Wiley, Chichester, 1973
2. D.J. Hartfiel. Dense sets of diagonalizable matrices. Proceedings of the American Mathematical Society, 123, 6, June 1995.
3. D.J. Hartfiel. Tracking in Matrix Systems. Linear Algebra and its applications, 165, 233-250, 1992.

Investigates the Attitude of Disabled Women towards the Childbirth in the City of Tehran

Ali Pezhhan and Parvaneh Afshary

Islamic Azad University, Tehran, Iran

This paper investigates the attitude of disabled women towards the childbirth in the city of Tehran using descriptive-survey method. The statistical population consists of all disabled married women in the city of Tehran (8481 women) according to 2011 census. Sampling was performed systematically and the sample size was 366 based on Cochran's formula, which finally reduced to 300. The data was collected using a researcher-made questionnaire consisting of 42 questions. Factor analysis results indicated that no correlation existed between items of one factor and those of another factor and that each factor independently affected the attitude of disabled women towards the childbirth. Chi-Square test results ($P<0.01$) indicated that a significant relationship existed between the attitude of disabled women toward childbirth and economic, social and cultural factors. According to demographic information results and F value ($P<0.01$), there was a significant difference between the attitude of disabled women in terms of occupation of husbands and social dignity and status of women. According to economic information and F value ($P<0.01$), there was a significant difference between the attitude of disabled women in terms of income level, family wealth and occupation. Cultural information results and Chi-Square test ($P<0.01$) indicated that a significant relationship

existed between the attitude of disabled women towards the childbirth and social welfare, environmental justice and the enjoyment of social security. Regression analysis results indicated that demographic factors (57%), social factor (69%), economic factor (58%), and cultural factors (63%) affected the attitude of disabled women towards the childbirth.
Keywords: Attitude, Disability, Disabled Women and Childbearing.

Bayes and Krige: Generalities

Giovanni Pistone[1], Grazia Vicario[2]

[1]de Castro Statistics, Collegio Carlo Alberto, Italy,
[2]Dept of Mathematical Sciences, Politecnico di Torino, Italy

In this paper we focus on the use of Kriging models for predicting at untried points from the response values at the tried locations, see e.g. [2]. The underlying Gaussian model requires the modelization of the covariance structure. In a previous paper [3] we have discussed to these aim generalities about the use of variograms to parameterize Gaussian models. In fact, Geostatisticians, pioneers and practitioners of Kriging, strongly support the variogram considering it more informative of the correlation structure. In particular computations for the case of jointly Gaussian $Y_1; \ldots; Y_n$ with constant variance $\sigma^2 = \mathrm{Var}(Y_i)$, $i = 1; \ldots; n$ are performed.
In such a case, the model can be parameterized by the common variance σ^2 and the variogram matrix □ that carries $n(n-1)=2$ degrees of freedom and is conditionally negative definite. The set of variogram matrices is related with a convex set called elliptope (ellipsoid+polytope), see [1]. The discussion of the domain for the variogram matrices is instrumental when viewing the problem in Bayesian terms. Opposite to the conventional geostatistical Kriging approach that commonly ignores the effect of the uncertainty in the covariance structure on subsequent predictions, a Bayesian approach will provide a general methodology for taking into account the uncertainty about parameters on subsequent predictions. Hence a-priori on the variogram matrices is demanded. We plan to discuss a number of simple typical problems with parameterized subset of variogram matrices and small dimension.
Keywords: Kriging, Variogram, Elliptope, Bayes.

References

1. Monique Laurent and Svatopluk Poljak. On a positive semidefinite relaxation of the cut polytope. Linear Algebra Appl., 223/224:439-461, 1995. Special issue honoring Miroslav Fiedler and Vlastimil Ptak.
2. Giovanni Pistone and Grazia Vicario. Kriging prediction from a circular grid: application to wafer diffusion. Appl. Stochastic Models Bus. Ind., 29(4):350-361, July 2013.
3. **Giovanni Pistone and Grazia Vicario. A note on (semi)variogram. In press. arXiv:1503.07686, march 2015.**

Spatial Modelling with Trends

P. M. Robinson
London School of Economics, U.K.

Models for spatial data recorded over time are considered. A wide range of spatial models are covered by our nonparametric treatment. Individual effects are allowed. The trends are of a general nature, allowing for both stationarity and nonstationarity.
Keywords: Time series, spatial data, nonparametric modelling, semiparametric modeling

Modelling Economic Count Data by Gaussian Distributions with Complex Parameters

Jose Rodríguez-Avi, Maria Jose Olmo-Jimenez
Dept of Statistics and Operational Research, University of Jaén, Spain

In several cases, we observe count data variables where the modal value is not zero, but another integer value. For instance, in the case of the number of facilities by municipality, we can see different behaviour depending on if the studied facility is a public or a private one (schools, health centres, libraries, etc.). In such cases data have overdipersion, but it may be due to a frequency of zeros higher than the Poisson – private ones - or a higher frequency of small values greater than 0. Even, for public equipments we may observe underdispersion, which is unusual in private cases.
In the case of overdispersion due to a high frequency of zero, usual count data discrete models are Negative binomial, Generalized Waring, Generalized Poisson distributions and so on. But in the underdispersed case, these distributions do not fit adequately. We propose the use of the CBP and CTP distributions, which belong to the family of Gaussian discrete distributions with complex parameters (Rodriguez-Avi et al, 2003, 2004, 2015), to model this type of data. In consequence, we present these distributions together with their properties and we use them to model several public and private municipal facilities in the Autonomous Region of Andalusia (Spain). In each case we compare the fits obtained by these distributions with those provided by other usual discrete distributions, showing the appropriateness of complex distributions for modelling public facilities.
Keywords: Count Data Variables, CTP distribution, Godness of fit

References

1. Rodríguez-Avi, J., Conde-Sánchez, A. and Sáez-Castillo, A. J., A new class of discrete distributions with complex parameters, Statistical Papers, 44, 67-88 (2003)

2. Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. J. and Olmo-Jiménez, M. J., A triparametric discrete distribution with complex parameters, Statistical Papers, 45(1), 81-95 (2004)
3. Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. J. and Olmo-Jiménez, M. J. Estimation and inference in complex triparametric pearson distributions, Communications in Statistics - Theory and Methods, 34(2), 277-289 (2005)
4. Rodriguez-Avi, J., and Olmo-Jimenez, M. J. A regression model for overdispersed data without too many zeros. Statistical Papers, DOI 10.1007/s00362-015-0724-9

Turnover Analysis in Manpower Planning

Evy Rombaut, Marie-Anne Guerry

Dept of Business Operations and Technology, Faculty of Economic and Social Sciences & Solvay Business school, Vrije Universiteit Brussel, Belgium

The ability to predict voluntary turnover is crucial for every company to prevent a broad range of financial losses. This prediction is usually based on time-consuming surveys with low response rates. The research question of this paper is whether the available data in the HR system can result in reliable predictions.
We will discuss the usage of decision trees and logistic regression. Each technique will require different data preparation and will bring different determinants of voluntary turnover to the surface.
An illustration will be given on a real life dataset of our own university. The model performance will be evaluated by the AUC measure, making it clear that the available data can be used by HR managers to prevent and reduce voluntary turnover.
Keywords: Human resources planning, Decision support system, Wastage, Turnover.

References

1. Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. (1984). Classification and regression trees. CRC press.
2. Hosmer, D. W., Lemeshow, S. (2000). Applied Logistic Regression, second ed. John Wiley & Sons, Inc.
3. Verbeke, W., Martens, D., Mues, C., Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications, 38(3), 2354-2364.
4. Wang, X., Wang, H., Zhang, L., Cao, X. (2011). Constructing a decision support system for management of employee turnover risk. Information Technology and Management, 12(2), 187-196.

Different Pathways in the Evolution of the Breast Cancer Using a Discrete Homogeneous Markov Model with Multiple Absorbing States

Ruiz-Castro[1], Juan Eloy and Zenga[2], Mariangela

[1]Dept of Statistics and Operations Research, University of Granada, Spain
[2]Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy

The multi-state models are usual in the survival literature to analyze the behavior of an illness passing through several stages. In this work, a discrete homogeneous Markov model with an indeterminate number of absorbing states is developed in an algorithmic and matrix form from a methodological point of view. Several absorption states are introduced in a general way, and a methodology that allows us to express the quantities in a well-structured form is applied. The model is applied for analyzing the behavior of breast cancer. A study of the relapse and survival times for 300 breast cancer patients submitted to mastectomy is presented.
Keywords: multi-state model, phase-type distribution, breast cancer, survival time.

A Mixed Effects Height-Diameter Model Based on Univariate Diffusion Process

Petras Rupšys[1,2], Edmundas Petrauskas[2]

[1]Centre of Mathematics, Physics and Information Technologies, Aleksandras Stulginskis University, Lithuania, [2]Institute of Forest Management and Wood Sciences, Aleksandras Stulginskis University, Lithuania

Statistical models using stochastic differential equations to describe dynamical evolution of natural systems are appearing in the scientific literature with some regularity in recent years. In this paper, our aim is to describe how stochastic differential equations can be used to aid our understanding of the height-diameter process of an individual tree. The tree height distribution was examined by using a Vasicek type stochastic differential equation and mixed effects parameters. The drift function depends on random effects and a diffusion term without random effects. The parameters were estimated by considering discrete samples of the diameter and height of individual trees and by using a maximum likelihood procedure. Used dataset was provided by the Lithuanian National Forest Inventory (LNFI) (2006-2010) from Scots pine trees. All results are implemented in a symbolic algebra system MAPLE.

Keywords: stochastic differential equation, maximum likelihood procedure, Vasicek's model, tree diameter, tree height.

Forecasting Mortality Trends Throughout a Three-Way Data Analysis

Giuseppe Giordano, Steven Haberman, Maria Russolillo

University of Salerno, Italy

The costs of the social security public systems, in almost all developed countries, are affected by two phenomena: an increasing survival in higher ages and a smaller number of birth. The combination of these two aspects largely impacts on the societies dealing with the rising pension and healthcare costs. In spite of the common trend given by the ageing population and the growing longevity, the mortality rates are also influenced by gender, countries, ethnicity, income, wealth, causes of death and so on. According to the WHO a “right” recognition of the causes of death is important for forecasting more accurately mortality. In this framework we intend to investigate the main causes of death impacting on the upcoming human survival, throughout a Multi-Dimensional Data Analysis approach to the Lee Carter model of mortality trends. In a previous paper, we stated that the crude mortality data can be considered according to several criteria. In this contribution we take into account a three-way array holding mortality data structured by time, age-group and causes of death. The model decomposition we propose is a modified version of the classical Lee Carter model allowing for three-way data treatment, analysis of residuals, graphical representation of the different components. A case study based on actual data will be discussed.

Measurement of Avoidable Mortality in Russia

Tamara Sabgayda, Victoria Semionova

Dept of analysis of health statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Russia

Russian mortality is characterized by a pronounced decline in the last decade. The rate of decline of avoidable and unavoidable mortality in population aged 5-64 years are close: in 2003 (the peak of mortality) until 2014, it decreased by 61.9% as against 59.3% in male mortality and by 52.1 against 48.1% in female death rate. For epidemiological monitoring in post-industrial European countries a new list of preventable causes was developed. In less developed countries the list which was

used during the last decade of last century is used yet. The question arises, whether the proximity of the rate of decline of avoidable and unavoidable mortality in Russia is due to the need to adjust the list of preventable causes to determine the actual indicator of health system.

To answer this question, we compared the effect of choice of list of preventable causes on the result of analysis in Russian Federation and the European Union in composition before May 2004.

List 1 (W. Holland, 1997) was adopted by experts for use in Russia in 2005. List 2 was proposed by E. Nolte and M. McKee (2004). In both lists, we grouped preventable causes into three levels of death prevention according to the approach of L. Simonato et al (1998). We used information from the European mortality database. As it represents a limited number of death causes the lists of preventable causes were not fully formed and the analysis was conducted for men and women aged 25-64 years for 1999 - 2013.

In 2013, avoidable mortality of men and women in Russia was higher than in EU at 4.1 and 2.2 times for List 1 and 6.1 and 3.2 times for List 2. From 2003 to 2013, mortality of men and women has decreased more in Russia than in EU: 1.7 and 1.6 times as against 23.8% and 13.1% for List 1 and 1.7 and 1.6 times as against 41.3% and 29.0% for List 2. The greatest reduction in Russian mortality comes in group 1 (death causes are preventable by primary prophylactic) for both sexes and both Lists. The greatest death reduction in EU comes in group 3 among men and in group 2 among women for List 1; in group 1 among both sexes for List 2, where group 1 accounts cerebrovascular disease only. In the EU, the death rate for Group 1 in the List 1 has changed little for ten-year period, but the rate of decline of mortality from causes of group 1 in the List 2 was the greatest. This means that the value of a healthy lifestyle is becoming understood by the overwhelming majority of European population. This is why the mortality from cerebrovascular disease depends on timeliness and quality of medical care only. In Russia, the change in mortality from cerebrovascular diseases occurs in parallel with change the rest preventable causes in List 1, making it more relevant to the current situation in the country.

Results of avoidable mortality analysis in EU depend significantly on list of preventable causes in contrast to Russia where such dependence observed for female mortality only. For EU, it isn't appropriate to include the causes that depend on primary prevention into the list of preventable causes. For Russia, new list of preventable causes can only serve as an additional source of information to the main results obtained on the basis of old list.

Keywords: avoidable mortality, list of preventable causes, indicator of effectiveness of public health system

American-Style Asian Options under Jump-Diffusion Processes with Arithmetic Average

Stefane Saize

Eduardo Mondlane University, Mozambique

Asian options are common financial derivatives in pricing currency market and commodity. These options reduce the risk of manipulation of the price at the maturity time. In a paper of Hansen and Jorgensen (2000) is studied the American-style Asian option with floating strike price. In this paper they consider the problem under diffusion processes. First of all, they transformed the problem into one-state variable problem and they established the general solution to the problem and its properties. In particular cases, they consider the problem when the average is geometric and arithmetic. In the arithmetic average case, the one-state variable is not a geometric Brownian motion, so they approximate it to a geometric Brownian motion using Wilkinson approximation. In a paper H. Pham (1997), is considered the standard American option under jump-diffusion processes is established the general solution to the problem. Also, H. Pham (1997) considered the case when the jump sizes are lognormal distributed. In a paper of C. R. Gukal (2001) is studied the standard American option under jump-diffusion processes using the Merton's idea.

In this paper we consider the American-style Asian option with floating strike, studied by Hansen and Jorgensen (2000), but in addition we consider under jump-diffusion processes.

From the book of S. Saize, we have the general solution for the American-style Asian options under jump-diffusion processes. So, we use this result to study the problem when the average is arithmetic. First of all we transform the problem into one-state variable as in Hansen and Jorgensen. To this new problem, we find its general analytical solution by using theories from Hansen and Jorgensen (2000), Merton (1976) and H. Pham (1997). In our case of study, when the average is arithmetic, the one-state variable is not a geometric Brownian motion, so we approximate it to a geometric Brownian motion as in Hansen and Jorgensen (2000) by using Wilkinson approximation and then we derive the solution to the problem. At the end, we find solution to the problem when the jump-size are lognormal distributed.

Keywords: American option, Asian option, Dual problem, Diffusion processes, Jump-diffusion processes, Equivalent measure, Lognormal jump sizes.

The Realized Quadratic Variation Derived from Trigonometric Transforms of Scaled Processes

Lino Sant

Dept of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

The statistical properties of realized quadratic variation, or more generally p-variation, have played important and crucial roles in solving problems involving hypothesis testing, diagnostic checking and parameter estimation for stochastic processes. Within the context of Levy processes, contrasts for corresponding sample statistics are exploited to infer about the continuity, or otherwise, of paths as well as about other matters involving special functionals and inherent properties like volatility. Realized variation can be split into parts originating from the canonical components of such processes. These considerations are many times extrapolated to the more general setting of Ito semimartingales.

In this paper a statistical functional involving the sine and cosine of a scaled version of a given process is defined. Its properties are studied and its performance using sample values is proposed as a general purpose method for investigating the nature of the generating process. Statistical results and simulation runs are proposed and discussed.

Keywords: Levy processes and semimartingales, realized quadratic variation, trigonometric transforms.

Assessment of Current Trends in Adult Mortality in The Context of Economic Changes in Russia

Victoria Semionova, Tamara Sabgayda

Dept of Analysis of Health Statistics, Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation, Russia

In last decade (2005-2014) Russia's GDP per capita increased by 2.4 times. At the same period, in parallel with that, the longest decreasing trends in mortality were observed in the country. Life expectancy increased by 6.4 years among men and by 4.1 years among women up to 65.3 and 76, 6 years respectively.

The aim of this study is to evaluate the changes occurred in the age and nosological aspects, as well as short-term forecasts of changes in life expectancy in the Russian economic context.

The increase in life expectancy in 2005-2014 was due to major causes and all age groups for which mortality has decreased at least a quarter. The highest rate of death decline was observed among population of 40-59 years old (39.6% and 33.8%). However, in last year (2013-2014) growth rates of life expectancy slowed sharply, reaching 0.1 years in both sexes (against the average annual growth of 0.6 and 0.4 years in 2005-2014). The rate of decline in mortality declined in all age groups of adult population by at least half. For the first time since 2005, negative changes were marked among working age population: 0.1% increase in mortality of men aged 0-39 years and 0.4% increase in mortality of 40-59 year old women (compared to the average annual decline in 2005-2014 which equal to 3.3% and 3.4% respectively). These negative changes were determined by 6.7% and 8.1% increase in mortality from digestive diseases, 1.5% and 4.1% increase in losses from infectious diseases. In addition, the mortality rate of 20-39-year-old man from diseases of the circulatory system grew by 1.4%, the mortality rate of 40-59 year old women from external causes grew by 2%. Even in those age groups where positive trends managed to keep, the increase was observed in mortality from certain classes. In particular, the growth of losses from diseases of digestive system is marked for all adults ages; the growth of mortality from external causes is marked for population aged 40 years or older. Thus, negative developments were determined firstly, by working age population, and secondly, by exogenous death causes.

It seems relevant question is whether this shift is fluctuation phenomena or the beginning of the formation of new stagnation or negative trend?

To answer this question, we note that in the last decade, Russia has experienced a similar slowdown in positive trends: in 2009-2010 the growth rate of life expectancy fell to ones observed in 2013-2014. Before it, the decline of GDP per capita more than a quarter was in 2008-2009. But by 2010, the GDP recovered, and with it has accelerated the positive trends of mortality. In 2013-2014, the rate of decline of GDP was less than half in 2008-2009 (12.1% vs. 26.4%), but the economic realities of modern Russia does not give reason to expect a quick exit from the crisis. There is no reason to assume a significant and sudden (multiple or close to that) decline in GDP per capita that results in almost simultaneous drop in life expectancy, as it was shown in 1990s in Russia. This suggests that the most likely scenario is stagnation in life expectancy. In this situation, the risk group will be working-age population that is traditionally for Russia. The loss will be generated primarily by exogenous death causes. It is safe to assume increased alcoholization of population as a dominant risk factor in Russia which will inevitably increase the importance of external death.

Keywords: short-term forecasts of changes in life expectancy, relation between GDP per capita and death rate, mortality in working-age population

Neural Networks Modelling for Estimation Total Waiting Time by Shortest Job First (SJF) Algorithm

Sibel Senan, Ruya Samli

Dept of Computer Engineering, Istanbul University, Turkey

CPU scheduling is an important subject in multiprocessing operating systems to satisfy maximizing CPU utilization which is one of the operating system goals. There are several scheduling algorithms (First-Come, First-Served scheduling, Shortest-Job-First scheduling, Priority scheduling, etc.) to execute processes and several characteristics (total waiting time, throughput number, turnaround time, response time etc.) to measure the efficiency of these algorithms. Hence, calculating a total waiting time of processes is one of the criteria for selecting a proper CPU scheduling policy. In this study, Neural Network (NN) approach is used for estimating the total waiting time of process sets which are first applied by Shortest Job First (SJF) algorithm. For this purpose, total waiting times are calculated for 40 process sets with randomly generated burst times of 10 processes whose arrival times are accepted as 0 by SJF policy. Then, a NN structure with back-propagation learning algorithm is designed for obtaining estimated results. Finally, these two result sets (calculated and estimated) are compared with each other to see if NN is an appropriate tool for modelling these implementations or not. This comparison shows that the difference between calculated and estimated results are less than 10% for all of the cases and less than 2.33% averagely which means the results of a designed NN model are highly accurate and using this model can be considered as an alternative and a successfull method to calculate waiting time of processes especially when the number of processes are too large to apply SJF algorithm is used.

Keywords: CPU scheduling, Neural Networks, Shorhest Job First Algorithm.

Risk-Neutral Densities Using General Entropy Measures in the Hunt-Devolder Semi-Markov Regime Switching Interest Rate Models

Muhammad Sheraz[1], Vasile Preda[2], Silvia Dedu[3]

[1]Dept of Mathematical Sciences & Dept of Economics and Finance, Institute of Business Administration, Karachi, Pakistan

[2]Bucharest University of Bucharest, Faculty of Mathematics and Computer Science and Romanian Academy, National Institute for Economic Research, Bucharest, Romania

[3]Dept of Applied Mathematics, Bucharest University of Economic Studies and School of Advanced Studies of the Romanian Academy, Bucharest, Romania

The application of entropy in finance can be regarded as the extension of both information theory and probability theory. Les Gulko applied Entropy Pricing Theory for pricing stock options and introduced an alternative framework of Black-Scholes model for pricing European stock options. In the field of econophysics, forecasting is a central problem and provides useful advanced techniques for the prediction of future movements. In the theory of option pricing the risk neutral probabilities play very important role. In this article we consider the Hunt-Devolder semi-Markov regime switching interest rate model, to find a solution of maximum entropy problem based on several general entropy measures. We obtain risk-neutral probabilities using Lambert function and a new type of approach.

Keywords: risk-neutral densities, entropy pricing theory, semi-Markov regime switching interest rate models, information measures.

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

Pricing Piecewise Barrier Options

Tomé Eduardo Sicuaio

Eduardo Mondlane University, Mozambique

The quality of the product gotten at the lowest price is something which financial instituitions, campanys and Goverments are looking for a sustantable growth and development.

The piecewise Barrier options inherite the standard barrier options failure. So the four main types of Piecewise barrier options are:

In the Up-and-out the spot price starts below the barrier level and has to move up for the option to be knocked out. Down-and-out, the spot price starts above the barrier level and has to move down for the option to become null and void. Up-and-in, the spot price starts below the barrier level and has to move up for the option to become activated. Down-and-in, the spot price starts above the barrier level and has to move down for the option to become activated.
Barrier options are some of the options used in many financial trading because of their lower price than standard vanilla options. This paper deals with the pricing of Piecewise Barrier options and compare it with the price of standard barrier options.
Keywords: Barrier Options, Piecewise Barrier Options, Knock out Options, Knock in options.

Tuberculosis in Brazil: New Cases, Healing and Abandonment in Relation to Level of Education in the Brazilian States

Edwirde Luiz Silva Camêlo, Dalila Camêlo Aguiar, Tânia M. R. Monteiro de Figueiredo, Andrés González Carmona, Ramón Gutiérrez Sánchez

Universidade Estadual da Paraíba, Brazil

SETTING: The situation of tuberculosis in Brazil, in each of its States and in the Federal District, with respect to patients' level of education.
OBJECTIVE: To analyse the strength of the relation between new cases of tuberculosis and treatment outcomes and rates of abandonment, with respect to the patients' level of education, in 2013.
DESIGN: Transversal-quantitative study, using data from the Brazilian Disease and Notification Information System related to notified cases of tuberculosis, taking into account the patients'level of education. Data were analysed by the hierarchical classification method (heat maps and dendrograms).
RESULTS: In relation to numbers of new cases of tuberculosis and rates of healing and treatment abandonment, the lower the patients'level of education (less than 9 years' formal education), the higher the numbers of new cases of tuberculosis and the higher the rates of healing and treatment abandonment, throughout the country. Levels of schooling of 1-5 years and of 6-9 years are especially strongly related to a high incidence of treatment abandonment, in all the States. This finding was confirmed by clustering techniques. Treatment abandonment is especially prevalent in the States of S~aoPaulo, Rio Grande do Sul and Rio de Janeiro, according to both analytic methods. In the States of

Minas Gerais, Esp'irito Santo, Mato Grasso do Sul and Parana, have the problem more similarly, but mixed with other groups of States.
CONCLUSION: Knowledge of the level of education of tuberculosis patients is of vital importance to health care managers, enabling them to strengthen healthcare policies aimed at effective control of the disease. This study highlights the situation in Brazil with respect to different levels of education. Programmes to eradicate tuberculosis should focus particularly on those patients who have had fewer years of schooling and thus are more likely to suffer this disease.

Least-Square Approximation Using Piecewise Power-Exponential Functions

Sergei Silvestrov, Karl Lundengård, Milica Rančić

Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Sweden

There are many processes that qualitatively can be described as an initial steep rise followed by a slower decay. One way to realize this behavior in a mathematical function is multiplying a power function with a shifted exponential function in a particular way. We will refer to this as a power-exponential function. This type of approximation has been used in several different circumstances, such as the modelling of fish behavior or the current in a lightning discharge. Here we demonstrate how construct approximations using linear combinations of piecewise power-exponential functions and how the presented framework allows for several rising and decaying parts. We will describe both least-squares fitting of the function to the measured data as well as fitting the first and second moments of the function using the Marquardt least-squares method and some applications.
Keywords: data fitting, piecewise power-exponential functions, Marquardt least-squares method.

Allied the Relative Risk with Initiation of Contraceptive Use in India and its Populous State Uttar Pradesh

Rohit Kumar Singh

Ministry of Health and Family Welfare, India

The paper attempts to examine the timing of initiation of contraceptive use at first time and after recent childbirth among currently married women and the relative risk associated with initiation of contraceptive

use by socio-economic and demographic characteristics. We tested the hypothesis-whether women who do not want any additional child initiate the contraceptive use early. Three rounds of NFHS data have been used. Cox-regression model has been used to analyze calendar data. The study reveals that a larger proportion of younger women start using a method without having any child. More than three-fourths women (15-19) begin to use family planning method with less than two surviving children, whereas most of the older women wait until they had at least two surviving children. Interestingly, for illiterate women the acceptance of family planning at 3+ living children as first use has gone up from 38 to 43 percent during 1992 to 2006. However, it is high among younger women. Prevalence of limiting method users is increasing over the period and most of the women have gone for sterilization in same month of last birth (i.e. around 35 percent) in India. The multivariate analysis suggests that programme variables like ANC and place of delivery (institution) affects the relative risk to initiate the use of contraceptive method after child birth.

Keywords: Family planning, contraception, limiting method

Parallel Stratified MCMC Sampling of AR-HMMs for Stochastic Time Series Prediction

I. Róbert Sipos

Dept of Networked Systems and Services,
Budapest University of Technology and Economics, Hungary

A Hidden Markov Model (HMM) (Baum and Petrie [2]) is a statistical model which is widely used for predictions in fields as diverse as finance and econometrics (Sipos and Levendovszky [10], Hassan and Nath [5], Mamon and Elliott [7], Hamilton [4]]), genomics (Durbin et al [3]), linguistics and speech recognition (Jurafsky and Martin [6], Rabiner [9]), and can model stochastic properties changing over time. One of its possible generalizations is the autoregressive-HMM (AR-HMM), which can describe an even wider range of random processes.

The model parameter estimation (learning) is a key aspect when we are using HMMs to predict future values of time series. Traditionally the Baum-Welch expectation maximization (Baum and Petrie [2]) algorithm is used, although it can provide only a local optimum. Unfortunately, computationally efficient algorithms have not yet been developed for finding a global maximum likelihood estimator of the free parameters. Moreover, when using HMMs for predicting future values (instead of predicting the most probable state sequence, as in many applications), a single maximum likelihood model estimation is not sufficient to capture

the stochastic behaviour. Especially, as they tend to be very sensitive for the observed time series, which in turn can be noisy.
Markov chain Monte Carlo (MCMC) (Andrieu et al [1]) is a method which applied in our case allows us sampling from the distribution of the HMM model parameters and therefore from the predictions by constructing a Markov chain with the same probability distribution. This sampling process makes the learning phase unnecessary, providing good quality results in a high speed manner.
The nature of the required algorithms to perform MCMC sampling of HMMs, including linear algebraic operations and the forward-backward algorithm (Rabiner [9]) for HMMs, exhibits such algorithmic properties that makes a massively parallel implementation feasible and beneficial. Besides the obvious speed-up, it inspired the use of stratified sampling yielding a better convergence. It also allows a generalization in the modelling via not having to set the number of hidden states in an ad-hoc way. In this light, the models are implemented using the concept of General-Purpose computing on Graphics Processing Units (GPGPU) to achieve superior performance compared to traditional CPUs both in terms of speed and power consumption (Nvidia [8]).
The performance of the proposed methods has been extensively tested on artificially generated data sets, and on real financial time series including equity indices, FOREX pairs, and commodity prices. The results demonstrate that a good predictive power can be achieved outperforming the traditional approaches. The stability of these methods in terms of sensitivity to the input sequence is also proven to be advantageous.

Keywords: Hidden Markov models, MCMC sampling, Stratified sampling, GPGPU, Financial time series

References

1. ANDRIEU, C., DE FREITAS, N., DOUCET, A., & JORDAN, M. I. An introduction to MCMC for machine learning. Machine learning, 50(1-2), 5-43, 2003.
2. BAUM, L. E. and PETRIE, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics 37 (6): 1554–1563, 1966.
3. DURBIN, R., EDDY, S. R., KROGH, A. and MITCHISON, G. Biological sequence analysis, Cambridge University Press, 1998.
4. HAMILTON, J. D. Analysis of time series subject to changes in regime. Journal of econometrics, 1990, 45.1: 39-70.
5. HASSAN, R. and NATH, B. Stock Market Forecasting Using Hidden Markov Model: A New Approach. Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications.
6. JURAFSKY, D. and MARTIN, J. H. Speech and Language Processing: An Introduction to Natural Language Processing. Computational Linguistics, and Speech Recognition, 2nd Edition, Pearson Prentice Hall, London, 2006.
7. MAMON, R. S. and ELLIOTT, R. J. Hidden Markov Models in Finance, Springer Science Business Media, 2007.
8. NVIDIA (2014) CUDA C programming guide [WWW]. Available from: http://docs.nvidia.com/cuda/cuda-c-programming-guide [Accessed 08/2014].
9. RABINER, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2), p. 257–286, February 1989.

10. SIPOS, I. R., LEVENDOVSZKY, J. High frequency trading with Hidden Markov Models by using clustering and PPCA algorithms. In: Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013), pp. 67-80. Barcelona, 2013.

A 4-Parameter Model for the Force of Mortality Based on a Second Order Approximation for the First Exit Time Densities

Christos H. Skiadas

ManLab, Technical University of Crete, Chania, Greece

In an earlier paper we have used a first order approximation for the first exit time densities of a Brownian motion through one-sided moving boundaries to model the mortality table data sets in Demography:

$$f_x = \left(\frac{2}{\sqrt{2\pi}}\right)\left(\frac{|H_x - xH_x'|}{\sqrt{x^3}}\right) e^{-\frac{H_x^2}{2x}} \quad (1)$$

H(*x*) is a function of health at age *x* and *f*(*x*) is the death probability density function.

We introduce the term $\left(2/\sqrt{2\pi}\right)$analogous to the related term of the half-normal distribution. This form has a relatively good fitting as is presented in figure 1 A, B. It is clear that the estimates in the age period from 15 to 50 years the model underestimates the related figures.

To improve fitting a second order approximation in the following form is proposed:

$$f_x = \left(\frac{2}{\sqrt{2\pi}}\right)\left(\frac{|H_x - xH_x'|}{\sqrt{x^3}} - \frac{k\sqrt{x^3}H_x''}{2|H_x - xH_x'|}\right) e^{-\frac{H_x^2}{2x}} \quad (2)$$

The parameter *k* expresses the level of the influence of the second order correction term. When *k*=0 the last equation form reduces to the first order approximation. The next step is to insert in (2) the 3-parameter (*b*, *l*, *c*) expression $H(x) = l - (bx)^c$ introduced by Skiadas and Skiadas [2] for *H*(*x*) to find the following form of model (3):

$$f_x = \left(\frac{2}{\sqrt{2\pi}}\right)\left(\frac{|l+(c-1)(bx)^c|}{\sqrt{x^3}} + \frac{k\sqrt{x^3}c(c-1)b^c x^{(c-2)}}{2|l+(c-1)(bx)^c|}\right) e^{-\frac{\left(l-(bx)^c\right)^2}{2x}} \quad (3)$$

This is a 4-parameter model providing quite well fitting for the death probability density data and for the logarithm of the force of mortality μ_x. We call (3) as Infant Mortality Model, IM-Model, giving quite good estimates for the infant mortality but also very good estimates for all the period of the life time as is illustrated in Figure 1 for USA males (A, C, F E) and females (B, D, F) in 2010.

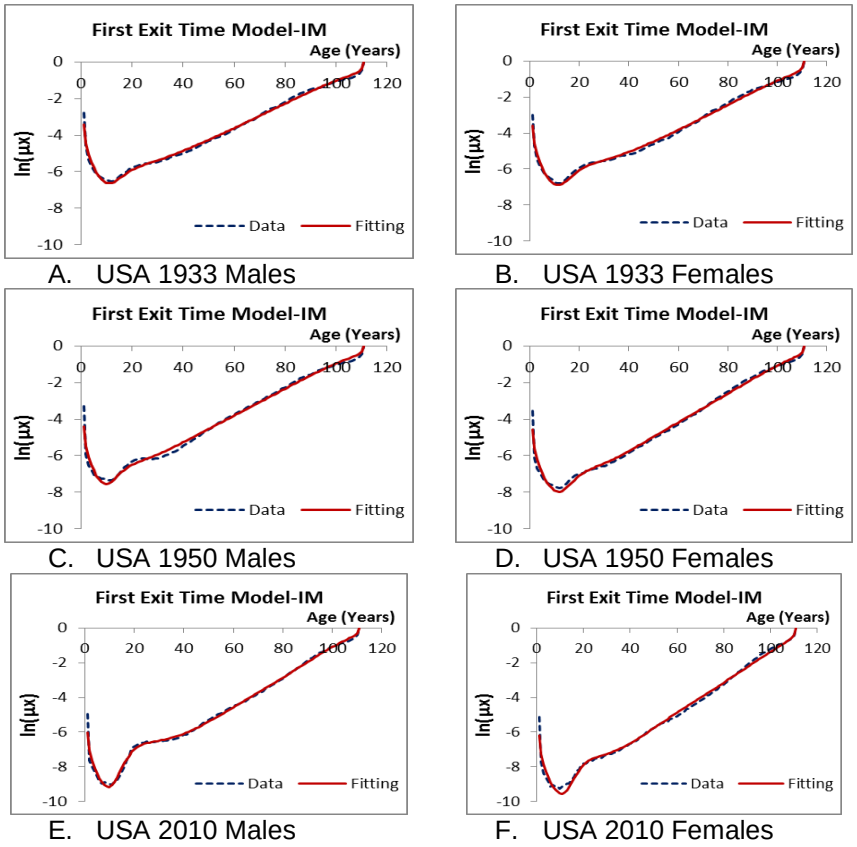

Fig. 1. The First Exit Time Model-IM including the Infant Mortality applied in the USA force of mortality data in logarithmic form for males and females the years 1933, 1950 and 2010.

This 4-parameter IM-Model is quite simpler than the SKI-6 parameter model or the 8-parameter model proposed by Heligman and Pollard. Further to the fitting ability of this model, more important is the demonstration that the first exit time theory used to model the mortality phenomenon is a very good tool both form the theoretical and applied point of view.

References

1. Skiadas, C. and Skiadas, C.H. Development, Simulation and Application of First Exit Time Densities to Life Table Data, *Communications in Statistics - Theory and Methods*, 39, 3, 444-451, 2010.

2. Skiadas, C.H. and Skiadas, C. 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*, 17, 973-982, 2015 (published online: June 2014).

Estimating the HALE Measurements of the Global Burden of Disease Study: Quantitative Methods and Applications in USA and Japan

Christos H. Skiadas

ManLab, Technical University of Crete, Greece

We propose a very simple model based on the mortality μ_x of a population provided in a classical life table and a mortality diagram to verify the Global Burden of Disease Study and the provided healthy life expectancy (HALE) estimates from the World Health Organization (WHO). Our estimates are compared with the HALE estimates for the World territories and the WHO countries. Even more we have developed the related simple program in Excel which provides immediately the Life Expectancy, the Loss of Healthy Life Years and the Healthy Life Expectancy estimate. We also apply the health state function theory to have more estimates and comparisons. The results suggest improved WHO estimates in recent years for the majority of the cases. See the related Figure for USA (males). The recently accepted estimates from WHO presented by red circles are within our range of estimates and the intervals set.

Keywords: Health state function, Healthy life expectancy, Mortality Diagram, Loss of healthy years, LHLY, HALE, DALE, World Health Organization, WHO, Global burden of Disease, Health status.

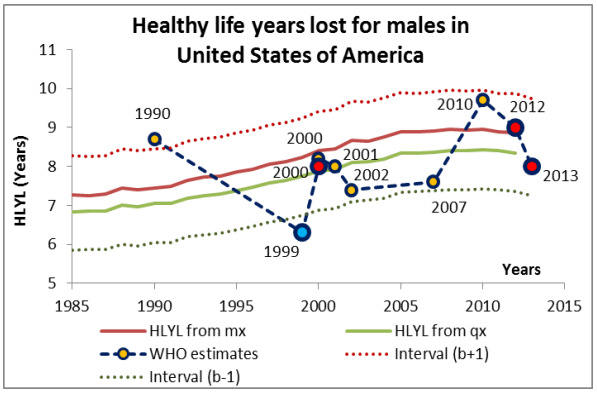

Fig. 1. Estimates of the Healthy Life Years Lost for USA (males)

A Method Based on the Chapman Kolmogorov Equation to Approximate Conditioned Diffusions

David Suda
University of Malta, Malta

Finding good proposals for simulating latent variables between the observations of a target diffusion process is important within the context of computational Bayesian inference. Computationally efficient proposals for diffusions which have been proposed tend to perform weakly with some target diffusions, particularly when observations are sparse. In certain cases, one would want to simulate from proposals which are as close as possible to the conditioned target diffusion, the latter requiring knowledge of the (frequently unknown) transition density. One approach to devising a proposal for target diffusions is that of approximating the transition density, and in this study we look at one numerical method based on the Chapman Kolmogorov equation, implementing it to examples of target diffusions which are difficult to emulate through efficient proposals. We show that, given the right time- and grid-discretisation settings, this proposal serves as a good approximation to simulate from, within reasonable computational time. We also compare the accuracy of importance sampling using this method at evaluating the transition density, in comparison to a numerical approach under identical settings.
Keywords: Diffusion, Importance Sampling, Bayesian inference

Rate of Convergence in Fuzzy non Homogeneous Markov Systems

Maria Symeonaki
Dept of Social Policy, School of Social and Political Sciences, Panteion University of Social and Political Sciences, Greece

Certain aspects of convergence rate in a non homogeneous Markov systems (NHMS) using fuzzy set theory and fuzzy reasoning are presented in this paper. More specifically, a fuzzy non homogeneous Markov system is considered where fuzzy inference systems are used to estimate the transition probabilities, the input and loss probabilities and the population structure of the system. It is proved that under some conditions easily met in practice, the rate of convergence of the sequence of the relative population structures in such a system is geometric, i.e. it converges to its limit geometrically fast.

Keywords: Markov systems, fuzzy set theory, convergence rate, geometric rate.

Gender-Based Differences on the Impact of the Economic Crisis on Labour Market Flows in Southern Europe

Maria Symeonaki, Maria Karamessini, Glykeria Stamatopoulou

Dept of Social Policy, School of Social and Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece

The present paper presents an application of the theory of non-homogeneous Markov system theory to labour force participation providing a cross-national, gender-based comparison of labour market transitions among Southern European countries. Raw data from the European Union Labour Force Survey (EU-LFS) from 2006 to 2013 is drawn, to compare the distribution of transition probabilities from the labour market state of employment, unemployment and inactivity and vice versa, for Greece, Italy, Spain and Portugal and examine whether the distributions are gender sensitive. Moreover, the paper examines whether school-to-work transition probabilities for these countries differ for males and females and to what extend. Additionally, the crisis' impact on the individual's labour market position is studied by country and by sex with the use of mobility indices.

Keywords: gender differences, labour market transitions, Markov systems, transition probabilities, EU-LFS

Measuring School-to-Work Transition Probabilities in Europe with Evidence from the EU-SILC

Maria Symeonaki , Maria Karamessini, Glykeria Stamatopoulou

Dept of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Greece

In the present paper the estimation of youth transition probabilities from school to employment is provided for all European countries using raw data drawn from the European Union's Survey on Living and Income Conditions (EU-SILC). More precisely, the input probabilities, i.e. the school-to-work probabilities are measured, using the EU-SILC data for the years 2006-2014. Replacing the European Panel Survey since 2003, EU-SILC provides timely and comparable cross-sectional, multidimensional on income, poverty, social exclusion and living conditions anchored in the European Statistical System. The theory of non-homogenous Markov systems (NHMS) will be used to measure the

school-to-work probabilities. In the proposed model a population is stratified into distinct categories according to a certain characteristic, which in our case is the labour market statuses. The study will deliver whether the crisis has created convergence or divergence in youth transitions to employment among European countries and study the evolution of school-to-work probabilities before and during the crisis for all European countries.

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).*

Estimating the Representativeness of German Parties and of the Bundestag Using the VAA Data

Andranik Tangian

Institute of Economic and Social Research in the Hans Böckler Foundation, Germany

The positions of German parties, Bundestag and Bundestag coalitions on 36 policy issues as given in the German voting advice application (VAA), the internet site *Wahl-O-Mat*, are compared with the results of public opinion polls, and the indices of popularity (the average percentage of the population represented) and universality (frequency in representing a majority) are constructed. The 2013 election winner, the CDU/CSU, is shown to be the least representative among the 28 parties considered. The most representative among the four parties in the Bundestag (with >5% of the votes) is DIE LINKE, which received only 8.6% of the votes. The representativeness of the Bundestag is about 50%, as if every policy were decided by coin tossing. To improve the situation, an alternative election method is discussed and hypothetically applied to redistribute the Bundestag seats with a gain in its representativeness.

Keywords: Mathematical theory of democracy; German parties; Bundestag election 2013; indices of representativeness

Visualizing Political Spectrum of Germany by Contiguously Ordering the Party Policy Profiles

Andranik Tangian

Institute of Economic and Social Research in the Hans Böckler Foundation, Germany

The data from the German voting advice application (VAA), the Wahl-O-Mat, are used to empirically construct and visualize the political spectrum of Germany. For this purpose, we consider the positions of 28 German parties on 38 policy issues declared before the 2013 federal election and associate the parties with the 38-dimensional vectors of their policy profiles. These vectors are used to define the party proximity in the party space. Principal component analysis (PCA) reveals that the parties constitute a thin ellipsoid whose two longest diameters cover 83.4% of the total variance. Reducing the model to just these two dimensions, a one-dimensional party ordering is found which is exactly the left–right axis rolled into a circumference, reflecting that the far-left and far-right ends of the political spectrum approach each other, although remain distant, resulting in its horseshoe-like shape. For comparison, contiguous party orderings are constructed using four other methods, and the results are discussed.
Keywords: German political spectrum, Voting advice applications (VAA), Contiguous ordering of variables, Principal component analysis.

Quantitative Trends in Engineering Education Media

Dvora Toledano–Kitai, Zeev Volkovich, Katerina Korenblat

ORT Braude College, Israel

Educational media is a kind of expression whose chronological configuration that mirrors through matter and style, evolution of the development in Engineering Education trends. In this work, we present a new approach, which makes possible to quantify tendencies in Educational media as a reflection of the developed educational approaches and methods in Engineering Education. We reveal sequential localization amid manuscripts via the study of the similarity structure in time series constructed for the considered media and characterize an accelerating rate of influence among those publications. Within a given timeslot, we quantify a content consistency with a specified subject, which makes possible to predict further directions of the educational methodology in Engineering Education.
Keywords: Engineering education, media quantization, text-mining

Robust Inference with Minimum Dual Divergence Estimators for Moment Condition Models

Aida Toma[1], Amor Keziou[2]

[1]Dept of Applied Mathematics, Bucharest Academy of Economic Studies, Piata Romana 6, and ISMMA, Romanian Academy, Romania, [2]Laboratoire de Mathematiques de Reims, Universite de Reims, Champagne Ardenne, UFR Sciences, France

The minimum dual divergence estimators and tests for moment condition models have been proposed recently in literature. The main advantage, of using a divergence based approach and duality, lays in the fact that it leads to asymptotic properties of the estimators and test statistics under the model, as well as under the alternative including misspecification, which cannot be achieved through the classical empirical likelihood context. Also, the estimators are asymptotically the best in the sense on Hansen yielding a "smallest" asymptotic covariance matrix. On the other hand, the estimators have bounded influence functions if and only if the function inducing the orthogonality conditions of the model is bounded. Since in many applications this function is not bounded, it is useful to have procedures that modify the orthogonality conditions in order to obtain robust versions of the estimators. In this paper we propose robust versions of the minimum dual divergence estimators using truncated orthogonality functions and duality techniques. We prove robustness properties and asymptotic properties of the new estimation method, underlying some advantages of using it with respect to other known methods.

Keywords: Moment condition models, robust estimation.

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

Algebra of Exploratory Factor Analysis with more Variables than Observations

Nickolay T. Trendafilov[1], Sara Fontanella[1], Kohei Adachi[2]

[1]Dept of Mathematics and Statistics, Open University, UK, [2]Graduate School of Human Sciences, Osaka University, Japan

The well-known exploratory factor analysis (EFA) model is defined for data with more observations than variables. This is dictated by the sizes and the structure of the matrix unknowns representing the EFA parameters. Nowadays, the most interesting applications require dimension reduction of data with much more variables than

observations. Thus, the standard EFA is out of date and impossible to apply.
The proposed work modifies EFA model to make it applicable to modern data with more variables than observations. This new development gives simple quantitative explanation of the notorious factor (scores) indeterminacy problem in classic EFA. The new EFA algorithm is very efficient in a sense it works with parameter matrices which dimensions are aligned to their ranks to avoid redundant calculations. The algorithm is illustrated numerically with a well know examples.
Keywords: Rank inequalities, Procrustes problems, alternating iterations.

Using the Common Principal Component Model to Estimate Covariance Matrices with Different Covariance Structures

Daniel Uys
Stellenbosch University, South Africa

Two or more multivariate normal densities with different covariance structures are considered. Although the covariance matrices are different, some or even all of their respective eigenvectors may be similar and therefore regarded as common. In situations like these, depending on the number of common eigenvectors, the common or partial common principal component model may be appropriate. Assuming the common principal component model, the covariance matrices share thus the same eigenvector structure, but with different sets of eigenvalues. The assumption of such a model can be helpful in the estimation of matrices with different covariance structures.
Biased estimators for the covariance matrices, based on the common or partial common principal component model, are proposed. The use of these biased estimators is illustrated within a multivariate setting.
Keywords: Covariance matrix, Common principal component model, Eigenvector, Eigenvalue

Latin Squares and Related Structures in Design of Experiment: A Statistical Perspective

Petya Valcheva[1], Teresa A. Oliveira[2,3]
[1]Sofia University "St. Kliment Ohridski",
[2]Departamento de Ciências e Tecnologia, Universidade Aberta, Portugal,
[3]Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal

Design of experiment can be considered at the point of greatest leverage of the Applied Mathematics and Statistics in many real-world situations. This new science is skillfully applied in diverse elds such as engineering, different kinds of medical experiments, agriculture and others.
Keywords: Design of Experiments (DOE), Latin Square(LS), Replicated Latin Square(RLS), Block Designs(BDs).

Representation of SW-Estimators

Jan Ámos Víšek
Institute of Economic Studies, Charles University, Czech Republic

The paper studies *S-weighted estimator* - a combination of *S-estimator* and *the Least Weighted Squares*. The estimator allows to adjust the properties, namely the level of robustness of estimator in question to the processed data better than the *S*-estimator or the Least Weighted Squares can do it. The paper offers the proof of its $\sqrt{n}-$ ***consistency*** and derives the asymptotic representation. Some pattern of results of the simulation study are also included, revealing the topology of data which shows that sometimes the outliers may be more dangerous than the leverage points.
Keywords: Robustness, implicit weighting, the order statistics of the absolute values of residuals, the consistency of the SW-estimator under heteroscedasticity, asymptotic representation.

Healthy Life Years: Which Age Groups Contribute Most and what Has Changed in Recent Years?

Jana Vrabcova Langhamrová [1], Jitka Langhamrová [2], Tomas Fiala[3], Jindrich Klufa[4]

[1]Dept of Statistics, Faculty of Informatics and Statistics, University of Economics, Czech Republic, [2]Dept of Demography, Faculty of Informatics and Statistics, University of Economics, Czech Republic, [3]Dept of Demography, Faculty of Informatics and Statistics, University of Economics, Czech Republic, [4]Dept of Mathematics, Faculty of Informatics and Statistics, University of Economics, Czech Republic

Demographic aging poses many challenges, but also risks. People in developed societies live longer, with increasing life expectancy. To know whether the extra years are improvement of health or increasing disabilities and dependencies is important as information for governments, health services, but also for individuals. Monitoring of

Healthy Life Years helps countries and regions to assess the health of their populations, to determine whether a longer lifetime spent in good health (compression of morbidity) or in bad health (expansion of morbidity). Healthy Life Years is used to highlight inequalities in health within countries, to allocate resources for health promotion, and they are increasingly used to gather the information needed for long-term planning of the health, social and fiscal policy.
To assess the development of mortality is most commonly used life expectancy of exactly x-year-old person. With decomposition method will be for selected years analysed contributions of age groups to the difference in life expectancy between selected years to highlight the development in time. Further will be calculated contributions of age groups to the difference healthy life years between women and men in selected years.
The aim of the paper will be to show how mortality rates are improving over the time, and gradually are improving mortality rates in middle-aged groups and shift to higher age groups.
Keywords: Healthy life years, life expectancy, sex, age, decomposition, Czech Republic, Malta

Index of Financial Pressure and Financial Crisis Forecasting (Fluid Approach)

Ognjen Vukovic
Belgrade Banking Academy, Serbia

The following paper tries to analyse financial crisis and contagion in a novel manner. It tries to prove that the future of finance is not based on Black-Scholes equation. It is the other equation that plays its role and that is, according to the author, much more important. The aforementioned equation is Navier-Stokes equation. Although Navier-Stokes equation is based and applied in fluid dynamics, its usage in finance is really innovative and interesting. After author of this paper defines and sets the foundation for Navier-Stokes equation application in fluid dynamics, index of financial pressure is introduced as a measure of forecasting financial crisis. Reynolds numbers for complexity of financial crisis and complexity of cash flows are introduced. At the same time, the author of this paper supports and introduces Mandelbrot fractal market hypothesis and provides evidence why it is superior to Fama EMH (ef! ficient market hypothesis). After having set the foundations for FMH hypothesis and Navier-Stokes equation application in advanced finance, author proves that the following assumptions are valid and have solid ground in predicting the financial crisis and financial contagion. The aforementioned approach represents ground breaking approach to

mathematical analysis of financial markets and provides a new look onto mathematical economics and finance.

An Intervention Analysis Based on the GARCH Model

Norio Watanabe, Masaki Nagashima
Chuo University, Japan

When an economical shock like the Lehman Crash occurred, it is expected to investigate the state of influence based on economical time series. The intervention analysis based on the ARIMA model by Box and Tiao is a method for such a purpose. However, there are few studies of intervention analysis for volatility of returns, which are typical eoconomical time series.
We propose a GARCH model for intervention analysis, since GARCH models are commonly used for analyzing returns of average stock prices or stock indices. Intervention terms in our model are considered for the mean value function and the volatility simultaneously.
The usability of the proposed model is demonstrated by applying to real time series.
Keywords: volatility, stock returns, intervention

A Genetic Algorithm Approach to Improve the Conservation of Centralities in Reverse Engineered Transcriptional Regulatory Networks

Holger Weishaupt, Patrik Johansson, Christopher Engström, Sergei Silvestrov, Sven Nelander, Fredrik J Swartling
Uppsala University, Mälardalen Högskola, Sweden

The identification of so called cancer driver genes, i.e. genes who's aberrant regulation in the cell is ultimately responsible for tumour development, represents one of the major foci of current cancer research. Among the techniques developed and employed for this task, a large body of methodology is dedicated to graph theory related methods that prioritize candidate genes based on their role in biological interaction networks. Specifically, these approaches often rely on the reverse engineering of transcriptional regulatory networks from gene expression data followed by the prediction of cancer genes based on network measures such as graph centralities or proximity to already known cancer genes. However, due to noise inherent to expression data and uncertainties in the exact nature of gene interactions, the reconstruction of accurate regulatory networks from gene expression

data has proven to be a particular challenging problem. The recent decade has spawned a multitude of methods specifically tailored to the task of reverse engineering of transcriptional regulatory networks. While these methods have proven ever more successful at accurately reconstructing individual biological interactions, we and others have recently shown that the systematic errors made by these techniques in combination with still limited numbers of correctly predicted edges lead to a loss of some of the global topological properties of the inferred networks. Specifically, such reconstructed networks might not preserve the gene centralities necessary for reliable cancer gene prediction. In this paper we investigate the possibility of combining a multitude of different network inference methods, exploiting the fact that these methods create varying networks with different systematic errors. Specifically, we are using a set of reference networks with corresponding expression data, from which reverse engineered networks are then reconstructed using a set of related methods. Subsequently, a genetic algorithm is then employed to obtain a combination of edges from the different inference methods, which is optimal with respect to the conservation of network centralities.
Keywords: Transcriptional network inference, simulated expression, graph centrality, genetic algorithm

Evaluating Performances of Different Multivariate Normality Tests by Using Multidimensional Scaling Technique

Soner Yiğit, Mehmet Mendeş

Çanakkale Onsekiz Mart University, Biometry and Genetics Unit, Turkey

This simulation study has been conducted to evaluate the performances of six different multivariate normality tests namely: Mardia's Skewness (M-S), Royston (Roy), Srivastava-Hui (S-H), Doornik-Hansen Test (D-H), Henze-Zirkler Test (H-Z), and Mardia's Kent (M-K) under different experimental conditions and also classifying these tests on the basis of their performances by using Multidimensional Scaling Technique. Obtained results of 50,000 Monte Carlo Simulation showed the most reliable when the Royston (Roy), Srivastava-Hui (S-H), and Doornik-Hansen test (D-H) have been applied. The above mentioned tests retained Type I error rates at nominal alpha level (0.05). Whereas, the estimations of Type I error of Mardia's Skewness (M-S), Mardia's Kent (M-K) and Henze and Zirkler (H-Z) test caused variations depending on sample size and number of variables. The estimations of test power of all tests have been affected by distribution shape, and the all related tests produced highly test power values especially when samples were taken from Multivariate Cauchy and Lognormal distributions. On the other

hand, the estimations of test power of all tests have been found extremely low when samples were taken from multivariate t-distribution with 10 d.f. Multidimensional Scaling (MDS) technique has been applied to classify the tests those have had similar performance and the factors those affected the performances of the above mentioned tests. At the end of Multidimensional Scaling analyses, it has been observed that the Roy, S-H and D-H tests showed similar performance, and the performances of these tests were obviously different than that of the others in general.
Keywords: Multivariate Normality, type I error, test power, multidimensional scaling technique

Multidimensional Scaling Analysis for Re-analyzing Results of Simulation Studies

Soner Yiğit, Hamit Mirtagioğlu, Mehmet Mendeş
Çanakkale Onsekiz Mart University, Biometry and Genetics Unit, Turkey

This study has been carried out to re-analyze the results of some simulation studies related to compare one-way ANOVA to its some parametric and non parametric counterparts. For this purpose, it is considered results of 25 studies published in journals that indexed by international indexes between 1980 and 2015. 25 studies have been grouped by using their results such as type I error rate, test power, variance ratio, total number of observation, sample size ratio, skewness and kurtosis. Experimental conditions have also been grouped in order to determine the factors that play important role in creating differences among 25 studies. Multidimensional Scaling Analysis results showed that the studies are divided into four groups and number of simulation, type I error rate and test power are the main factors that affect creating differences among the studies.
Keywords: Simulation, Multidimensional Scaling, Type I error rate, test power

A Simple and Effective Technique for Forecasting Mortality Patterns

K. Zafeiris
Laboratory of Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece

In population projections the problem of forecasting age-specific mortality trends is of central importance. In this paper a new technique serving this purpose is proposed. Given the death rates for wide age groups ($_nq_x$), the relational technique of Kostaki (2000) is applied for the estimation of one-year death probabilities (q_x). Then, a mixture of the nine-parameter version of Heligman-Pollard model (HP9) as developed Kostaki (1992) and two subsequent cubic splines were used. In that way the advantages of the HP9 method are preserved while the flexibility of the cubic splines allows the application of the method on various shapes of mortality patterns. After the age of 99 years the probabilities of death were extrapolated on the basis of the parameters of the last spline used. Results of this technique using empirical data indicate that the method proposed is very effective and also while quite parsimonious in terms of simplicity.

Recent Mortality Trends in Greece

K. N. 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 work examines the mortality characteristics of the population of Greece for the years 1961 to 2014. The time period considered is of special interest for the modern demographic history of the country, since it is characterized by enormous developments on economic, political and social grounds. For that, the life tables of the male and female population of the country were constructed and several mortality indicators like life expectancy at chosen ages were calculated. For the purpose of analysis, a statistical technique for smoothing the age-specific death rates in order to estimate the theoretical age-specific probabilities of dying is developed. This technique consists of a combination of the nine-parameter version of Heligman-Pollard formula proposed by Kostaki (1992) and cubic splines. The results of our calculations are indicative of the mortality transition observed in the country the last 55 years. There is also conclusive evidence for the appearance of the first effects of the economic crisis in the country.

A Method for the Evaluation of Health Trends in Greece, 1961-2014

K. N. Zafeiris[1], Christos 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 period 1961-2014 is of great importance in the Modern Greek History. In the 1960s several reformation efforts, which took place until 1965, were followed by a period of political turmoil and finally the dictatorship of the Colonels in April 1967. The country returned to democratic normality in 1974. In 1981 Greece became a full member of the European Union and until 2008 followed a rapid developmental course. Then, the economic crisis caused the rapid decrease of GDP and the aggravation of the socio-economic characteristics of the population because of the austerity policies which were applied. The scope of this paper is to analyze the health levels of the population of the country during that period with the application of a newly proposed method for the calculation of healthy life expectancy. Results indicate the rapid improvement of health of the Greek population; however, there is evidence of the effects of the economic crisis on population health.

On Maximum Likelihood Estimators and their Modifications: Multivariate Case

Aleksander Zaigrajew

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Poland

A multivariate problem of parameters' estimation in regular models is considered. As a class of estimators we consider that based on the maximum likelihood estimators. The main goal is to compare the estimators with respect to the second order risk based on the weighted mean squared error. Our approach is based on the notion of the second order admissibility of estimators and methodology given in [1]. In contradistinction with situation described in [3], here we consider the multivariate problem without nuisance parameters. The similar problem of estimating both shape and scale parameters of the gamma distribution is considered in [2].

Keywords: Maximum likelihood estimator, Weighted mean squared error, Second order risk, Admissibility.

References

1. J. K. Ghosh and Bimal K. Sinha.A necessary and sufficient condition for second order admissibility withapplications to Berkson's bioassay problem. Annals of Statistics, 9, 6, 1334-1338, 1981

2. H. Tanaka, N. Pal and W. K. Lim. On improved estimation of a gamma scale parameter. Statistics. Unpublished manuscript, 2014

3. A. Zaigrajew and I. Kaniovska. A note on comparison and improvement of estimators based on likelihood. Statistics, 50, 1, 219-232, 2016

On the Asymptotics of the Random Sums

Nadiia Zinchenko

Dept of Informatics and Applied Mathematics,
Nizhyn State Mukola Gogol University, Ukraine

Our work deals with the rate of growth of random sums D(t) = S(N(t)), where N(t) is a renewal counting process and S(t) is an usual partial sum process. We proposed a number of integral tests for investigation of the upper/lower functions for D(t) under various assumptions on renewal process, moment and dependent conditions of random summands $\{x_i, i \geq 1\}$. The cases of independent, weakly dependent and associated summands are studied as well as random variables satisfying φ-mixing conditions. Corresponding proofs are based on the rather general theorems about the strong approximation of the random sums by a Wiener or α-stable Lévy processes (Zinchenko (2015)). The same approach is used to study the asymptotic behavior of increments D(t+a(t))-D(t) on intervals, whose length a(t) grows, but not faster than t. As a consequence, various modifications of the LIL and Erdös-Rényi-Csörgő-Révész-type SLLN for random sums are obtained and used for investigation fluctuations of the risk processes in classical Cramer-Lundberg and renewal Sparre Andersen risk models. Finding the non-random boundaries f(t) for the rate of growth of D(t) and their increments helps to analyze and plan insurance activities and to determine the value of insurance premiums and reserves. The case of risk processes with stochastic premiums (Zinchenko (2014)) is also discussed.

Keywords: Random sums, Renewal process, Strong limit theorems, Strong Invariance Principle, Rate of growth, Law of iterated logarithm, Risk process.

References

1. N. Zinchenko. Almost sure approximation of the superposition of the random processes. Methodology and Computing in Applied Probability, 17, 1, 235-250, 2015

2. N. Zinchenko. Strong limit theorems for the risk processes with stochastic premiums. Markov Processes and Related Fields, 20, 3, 527-544, 2014

Modelling activity in an Australian Emergency Department using Coxian phase-type distributions to capture key patient pathways

Laura Boyle[1], Adele H Marshall[1] and Mark MacKay[2]
[1]*Centre for Statistical Sciences and Operational Research (CenSSOR)*
Queen's University Belfast, Belfast, Northern Ireland (UK)
[2]*Health Care Management, School of Medicine,*
Flinders University, Bedford Park, South Australia

Hospital emergency departments (EDs) are a key access point for individuals requiring urgent hospital care. In Australia, all public hospital EDs are subject to the National Emergency Access Target of treating or admitting patients within four hours, however the proportion of patients at South Australian EDs seen on time in 2014-2015 was 66%, falling 23% short of the national target (AIHW[1]). Development of accurate models for EDs can assist hospital service planners with analysing patient flow and planning strategic improvements of the system. Patient flow through EDs is complex due to the diverse range of patient ailments and varying patient resource requirements. The aim of this research is to use Coxian phase-type distributions to model the key patient pathways within a South Australian ED. Coxian phase-type distributions are a type of Markov model which represent the time until absorption of a finite Markov chain in continuous time as a series of latent phases (Neuts[3]). The characteristics of patients that progress to the later phases can be identified, revealing groups of individuals who consume a greater proportion of resources and are likely to exceed the four-hour target (Marshall and McClean[4]). Future work involves using the resultant Coxian fits to populate a simulation model of the ED.

Keywords: Coxian phase-type distributions, Markov models, Emergency care

References

1. Australian Institute of Health and Welfare 2015. Emergency department care 2014- 15: Australian hospital statistics. Health services series no. 65. Canberra: AIHW.
2. M.F. Neuts. Matrix-geometric solutions in stochastic models: an algorithmic approach, MIT Press, Cambridge, MA, 1981.
3. A.H. Marshall and S.I. McClean. Using Coxian Phase-Type Distributions to Identify Patient Characteristics for Duration of Stay in Hospital. Health Care Management Science, 28, 2, 79–93, 2005.

Stress Dependence and Enforced Regression Paradigm

Jerzy K. Filus[1] and Lidia Z. Filus[2]
[1]*Department of Mathematics and Computer Science*
Oakton Community College, Des Plaines, IL 60016, USA
[2]*Department of Mathematics*
Northeastern Illinois University, Chicago, IL 60625, USA

We describe a way human residual life time distribution depends on amounts of stresses the given person is subject to. Using our method of the parameter dependence (2013) we describe stochastic dependence of a random variable, say, Y from a set of explanatory variables $X_1, \ldots, X_k$ as conditional survival function $P(Y \geq y \mid x_1, \ldots, x_k)$. Relative easiness of the construction of these conditional distributions, when using the method of parameter dependence, prompted us to apply them as an alternative to the existing regression theory. This alternative relies on replacing the, typically used in the regression theory, conditional expectation, say, $E[Y \mid x_1, \ldots, x_k]$ by the more general notion of conditional probability distribution of Y, given the realizations $x_1, \ldots, x_k$. This new approach is expected to improve accuracy of the existing dependence models. It also essentially differs from the "Regression Quantiles" concept by Koenker and Basset (1978) since our method is parametric and results in analytic formulas for the probabilities $P(Y \geq y \mid x_1, \ldots, x_k)$ as continuous functions of the variable y.

References

1. J. K. Filus, L. Z. Filus, A Method for Multivariate Probability Distributions Construction via Parameter Dependence, *Communications in Statistics: Theory and Methods*, Vol. 42, Num. 4, 15, 716-721, 2013.
2. R. Koenker, G. W. Bassett, Regression Quantiles, *Econometrica*, 46, pp 33-50, 1978.

Applications in the healthcare domain for the quality of life via semi Markov modelling

Zacharias Kyritsis and Aleka Papadopoulou
Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54124,Greece

In the present paper a non homogeneous semi Markov reward model for a patient's quality of life is considered. The basic assumption of the model is that quality of life depends on its health status. The patient's expected interval reward and the variance are calculated by means of the basic parameters of the system. The above measures can be used

as an index of the quality of life status for all health states. Illustrations using simulated healthcare data are presented and the corresponding results are provided.
Keywords: Healthcare, Quality of Life, semi-markov model, rewards

Incorporating patient covariates into the conditional Coxian phase-type distribution to model elderly patient movement through hospital and community care

Adele H Marshall and Andrew S Gordon
Centre for Statistical Sciences and Operational Research (CenSSOR)
Queen's University Belfast, Belfast, Northern Ireland (UK)

This research presents a new methodology extending the conditional Coxian phase-type distribution to include covariates. The probability density function for this distribution is computationally expensive due to the multiplicative matrix elements contained therein. However, employing an analytic form for the general r-phase conditional Coxian phase-type distribution, the computational expense has been greatly reduced. The methodology has been applied to the field of elderly patient care and used to model the transition of elderly patients between hospital and community care. With the inclusion of patient covariates, for the rest time in the literature, the effect of various patient characteristics on the length of time spent in both hospital and community care may be discovered. A data set containing hospital readmission data for elderly patients is used an illustrative example.
Keywords: Survival analysis, conditional Coxian phase-type distribution, hospital readmissions

A Markov approach to discrete-time semi-Markov healthcare systems

Aleka Papadopoulou[1], Sally McClean[2] and Lalit Garg[3]
[1]Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54124,Greece
[2]School of Computing and Information Engineering, University of Ulster, Coleraine, Northern Ireland
[3]Computer Information Systems, Faculty of Information & Communication Technology, University of Malta, Malta

Previously a number of authors have used Markov models to study the movement of patients through hospital care where each patient spends an amount of time in hospital, referred to as length of stay (LOS) and is

then discharged, often to be replaced by another patient from the waiting list. Semi-Markov modelling can allow us to build more realism into the LOS distributions and provide more accurate predictions. However, the models easily become intractable and there are also difficulties concerned with obtaining detailed data to support parameter estimation. We here circumvent these difficulties by regarding each patient pathway as a state of the semi-Markov model; thus the holding time distribution of each state of the semi-Markov process is characterised as a LOS distribution for the corresponding patient pathway. By assuming a closed system we envisage a situation where the hospital system is running at capacity, so any discharges are immediately replaced by new admissions to hospital. In this paper we describe such a semi-Markov model and provide first and second moments of numbers in each semi-Markov patient pathway at any time and in the limit using Markov models based on daily transitions. Such key performance indicators are useful for future capacity planning for hospitals with a need to optimally allocate limited resources. Our model is illustrated using stroke data from a large hospital. Stroke patients typically are very consuming of resources and often require long periods of hospitalization. As such, they provide a good example for illustrating the use of such models.

Keywords: Healthcare Markov models, Semi Markov process, Patient planning.

Monte Carlo Methods Applied in Health Research

José A. Pereira1, Luzia Mendes2, Antonio Costa3, and Teresa Oliveira4

[1] Faculdade de Medicina Dentária da Universidade do Porto e MBB - Universidade Aberta, Portugal,
[2] Faculdade de Medicina Dentaria da Universidade do Porto, Portugal
[3] Universidad Nacional de Educacion a Distancia, Madrid, Spain
[4] CEAUL e Universidade Aberta, Lisboa, Portugal

The root surface area of the tooth (RSA) is an important prognostic factor in the field of dentistry. The estimation of RSA from routine clinical data, such as tooth length (TL) and mesiodistal diameter of crown (MDC), is of interest because provides clinicians with information to decide objectively without additional costs. The aim of the paper is to determine the sample size for a regression analysis of RSA on TL and MDC using both power and parameter accuracy perspectives with Monte Carlo (MC), as describe by Beaujean (2014).

A random sample of 5 lower second premolar teeth where scanned in X-ray microtomograph and their RSA where obtained through planimetric method, the TL and MDC measured on 1:1 photographs. The model of

interest was defined as $RSA=\beta_0+ \beta_1TL+ \beta_2MDC$, according with the research question. The sample size was determined based on the model of interest and strength of the relations among the variables using the MC methods. The packages *lavaan* and *simsem* of R software were used to define the model and to run the simulations.
The sample size that meets the criteria for Monte Carlo data quality proposed by Muthén & Muthén (2002) was calculated.
Keywords: Root surface area, sample size, Monte Carlo methods.

AUTHORS INDEX

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

Y

Z