SMTDA 2014

Stochastic Modeling Techniques and Data Analysis International Conference and Demographics Workshop Lisbon, Portugal, 11-14 June 2014

Plenary and Keynote Talks

On Cluster Analysis of Complex and Heterogeneous Data

Helena Bacelar-Nicolau[1], Fernando C. Nicolau[2], Áurea Sousa[3], Leonor Bacelar-Nicolau[4]

[1]Faculty of Psychology, LEAD; ISAMB, CEA; University of Lisbon, [2]FCT, Department of Mathematics, New University of Lisbon. [3]Dep. of Math., CEEAplA, CMATI, University of Azores. [4]Faculty of Medicine, Institute of Preventive Medicine, ISAMB, University of Lisbon, Portugal

Classical clustering methods usually work with a set of objects as statistical data units described by a set of homogeneous (that is, of the same type) variables in a two-way framework. This paradigm can be extended in such way that data units may be either simple / first-order elements (e.g., subjects, individuals from a basic population) or groups of / second-order or more objects from some population (e.g., subsets of the population, subsamples of a sample, classes of a partition) and/or descriptive variables may simultaneously be of different (e.g., binary, multi-valued, histogram or interval) types. Therefore, one has a complex and/or heterogeneous data set under analysis. In that case classification will often be carried out by using a three-way or a symbolic/complex approach.

The present work synthesizes previous methodological results and shows several developments mostly regarding hierarchical cluster analysis of three-way and of symbolic data, where statistical data units are described by either a homogeneous or a heterogeneous set of complex variables. We illustrate that approach on a few examples from the statistical literature. The methodology has been applied with success in data mining context, concerning multivariate analysis of real-life data bases issued from economy, medicine and social sciences.

Keywords: Cluster analysis, Similarity coefficient, Hierarchical clustering model, Three-way data, Symbolic data, Heterogeneous data.

Principled approaches to the interpretation of machine learning classifiers

Paulo Lisboa
Liverpool John Moores University, UK

Machine learning has often focused on matters of accuracy and computational efficiency. In so doing, it has become apparent that many approaches exist for flexible models with essentially equivalent performance once they are optimised to a particular application. Moreover, an estimate of a posterior probability, however accurate in principle, is likely to be of limited practical value unless its value can be understood by a domain user without specific expertise in data modelling.
The talk will argue that the value of probabilistic classifiers is contained in much more than its outputs, rather in the metric structure that it lends to data space. This makes the application of the machine learning method intuitive to the use and, more important, verifiable against domain expertise. Moreover, from a theoretical view point this approach to classification explicitly addresses the fundamental question of what is similarity? Using Fisher information to derive a Riemannian metric in data space, the notion of similarity can be rigorously quantified way and put to into practice for use by experts in the application domain.

Modelling Toroidal Data Using Circulas

Arthur Pewsey
University of Extremadura, Spain

We consider ‘circulas’; the circular analogues of copulas. More specifically, we concentrate on one particular class of bivariate circulas which is pre-existing but has not been studied in such explicit form or detail before. This class is appealing in many ways but does not necessarily result in especially attractive bivariate circular models when the marginals are not circular uniform. A major exception is an elegant bivariate wrapped Cauchy distribution proposed and developed by Kato & Pewsey (2013). We will consider properties of the circulas themselves, as well as those of distributions generated using circulas with marginals that are not circular uniform. Likelihood based inference for the latter distributions will be considered and applied in the modelling of wind directions at a Texan weather station and data on the pre-earthquake direction of steepest descent and post-earthquake direction of lateral

ground movement before and after, respectively, an earthquake in Noshiro, Japan.
Key Words: bivariate; circular distributions; copula; dependence; multivariate; torus

What is the relationship between longevity and health?

Jean-Marie Robine
INSERM (French National Institute of Health and Medical Research) and EPHE (Ecole pratique des hautes etudes), Paris and Montpellier, France

What is the relationship between longevity and health? Health expectancies were developed more than thirty years ago specifically to answer this question. It may therefore be the time to try to answer this question, though it is worth noting that the question implies a uni-directional relationship. Almost no-one questions the positive association between health and longevity. It is expected that healthy robust people will live on average longer than frail people. On the other hand many people wonder about the relationship between longevity and health. Are we living longer because we are in better health? Are we living longer in good health? Or are we merely surviving longer whatever our health status? In others words, can we live in good health as long as we can survive. And this is exactly the purpose of health expectancies: monitoring how long people live in various health statuses.
Firstly at an individual level, we address whether centenarians are healthy people. Secondly at the national level, we answer whether the increase in healthy life expectancy is slower or faster than the increase in total life expectancy. Finally at the global level we address whether countries with the highest life expectancies also have the highest healthy life expectancies.

Invited and Contributed Talks

Proximity measures in topological structure for discrimination

Rafik Abdesselam
COACTIS-ISH Management Sciences Laboratory - Human Sciences Institute, University Lumière of Lyon 2, France

In this paper, we introduce a new topological approach using proximity measures with a discrimination and classification aim.
In many application fields, the choice of a proximity measure directly affects the results of data mining methods in comparison or structuring a set of objects. Generally, we choose one proximity measure among many existing ones. However, some are more or less equivalent.
The proposed approach is based on topological equivalence which exploits the concept of local neighbors. It defines discriminant equivalence between two proximity measures as having the same neighborhood structure on the objects of a set of explanatory continuous variables according to a target qualitative variable that we want to explain.
According to the notion of topological equivalence based on the concept of neighborhood graphs, we use adjacency binary matrices, associated with proximity measure, between and within groups to classify. Some of the proximity measures are more or less equivalent, which means that they produce, more or less, the same discrimination results. It believes that two proximity measures are topologically equivalent if they induce the same neighborhood structure on the objects in purpose of discrimination and classification of anonymous objects.
The comparison adjacency matrix is a useful tool for measuring the degree of resemblance between two empirical proximity matrices in a discriminating context.
We first compare statistically the obtained confusion tables, then, to view these proximity measures, we propose an hierarchy of proximity measures which are grouped according to their degree of resemblance in a topological context of discrimination.
We illustrate and compare empirical results obtained on thirteen proximity measures on a real example data.
Key Words: proximity measure, discrimination and classification, dissimilarity and adjacency matrices, neighborhood graph, topological equivalence.

Univariate skew-unimodal distributions with mode-preserving property

Toshihiro Abe, Hironori Fujisawa
Tokyo University of Science, Japan

We consider an asymmetric family of unimodal distributions from a family of the symmetric distributions by using a specific class of transformation controlling the skewness. Its mode and normalizing constant remain unchanged for the transformation. Monotonicity of skewness can be shown easily under a weak condition for a transformation due to its mode invariance property. The application of the proposed model and inferential methods is illustrated in analyses of two data sets.
Key Words: Entropy maximization; Mode preserving; Monotonicity of skewness; Skew-symmetric distribution; Unimodality

Time Series Segmentation by Switching Markov Model based Fuzzy Clustering

Janos Abonyi
University of Pannonia, Department of Process Engineering, Hungary

Regime switching models are attractive when a nonlinear system should be modelled and simultaneously i) the data partition, ii) time instant of changes, and iii) parameters of local models should be determined. For the identification of these models multiple and operating region based modeling techniques, hinging hyperplanes, mixture of regression models, and hidden Markov models can be used.
We focus on the identification of time-varying dynamical systems and develop an easily implementable and interpretable identification algorithm based on fuzzy clustering. The resulted model based time-series segmentation algorithm can be considered as a constrained clustering: multivariate data points (objects) should be grouped by their similarity under the constraint of all points in a cluster must come from successive time instances, and changes of the membership degrees should be interpreted as a Markov process.
Although we already modified fuzzy clustering algorithms for time-series segmentation, to represent the dynamics behind the switching process further developments are needed. The probabilities of making a switch from one state to the other are controlled by the transition matrix that should also be identified in our novel model based fuzzy clustering algorithm.
The proposed algorithm is applied to the modeling and monitoring of multivariate stochastic systems. An application example related to the energy monitoring of an industrial process will be also presented.

An inverse first-passage problem for one-dimensional diffusions reflected between two boundaries

Mario Abundo
Dipartimento di Matematica, Università Tor Vergata, Italy

We study an inverse first-passage-time problem for a one-dimensional, time-homogeneous reflected diffusion $X(t)$, which is the solution of the stochastic differential equation with reflecting boundaries α and b:

$$dX(t)= \mu(X(t))\, dt + \sigma(X(t))\, dB(t) + dL(t) - dU(t), \quad X(0) = \eta \in [a,b],$$

where $B(t)$ is standard Brownian motion, the initial position η is a random variable, independent of $B(t)$, and the coefficients are sufficiently regular, so that there exists a unique strong solution $X(t)$ which remains in the interval $[a,b]$ for every time. The processes $L = \{L(t)\}$ and $U = \{U(t)\}$ are the regulators of points α and b, respectively; they are uniquely determined by the following properties (see Harrison (1985)):
(i) both $L(t)$ and $U(t)$ are continuous and non-decreasing with $L(0) = U(0) = 0$;
(ii) $X(t) \in [a,b]$ for every $t \geq 0$; (iii) L and U increase only when $X = a$ and $X = b$, respectively, that is, for $t \geq 0$:

$$\int_0^t I\{X(s) = a\}\, dL(s) = L(t) \quad \text{and} \quad \int_0^t I\{X(s) = a\}\, dU(s) = U(t).$$

Let be $S \in [a,b]$ a barrier such that $P(a \leq \eta \leq S) = 1$, we consider the FPT of X over S, i.e. $\tau_S = \inf\{t > 0 : X(t) = S\}$, and set $\tau_S(x) = \inf\{t > 0 : X(t) = S \mid \eta = x\}$. We assume that for every $x \in [a,S]$, $\tau_S(x)$ is finite with probability one and that it possesses a density $f(t|x)$. Usually, in the FPT problem the initial position is assumed to be deterministic and fixed to a value x, then the direct problem consists of finding the FPT density $f(t|x)$. Notice that, while for ordinary diffusions a certain number of results on the direct FPT problem is available in the literature, for reflected diffusion processes only a few papers appeared recently.

We study the following inverse first-passage-time (IFPT) problem. Let be $S \in [a,b]$, with $P(a \leq \eta \leq S) = 1$; for a given distribution F, our aim is:

to find the density g of η (if it exists) for which it results $P(\tau_S \leq t) = F(t)$

This IFPT problem has interesting applications in Mathematical Finance, in particular in credit risk modeling, where the FPT represents a default event of an obligor. For ordinary diffusions, it was studied by Jackson et Al. (2009) in the case of Brownian motion, while some extensions to more general processes were obtained by Abundo (2012, 2013).

References
1. Abundo, M., Stat. Probab. Letters, 82 (1), 7-14 (2012)
2. Abundo, M., Stat. Probab. Letters, 83, 168-176 (2013).
3. Harrison, M., Brownian motion and Stochastic Flow Systems. Wiley, New York (1985).
4. Jackson, K. et Al., Stat. Probab. Letters, 79, 2422-2428 (2009).

Different heuristic optimization techniques for traffic lights signal timing problem

Rami ka. Abushehab[1], Baker Kh. Abdalhaq[2], Badie Sartawi[1]
[1]Alquds university, [2]An-najah university, Palestine

A good controlling for the traffic lights on the network road may solve the traffic congestion in the cities. This paper deals with the optimization of traffic lights signal timing. We used four different heuristic optimization techniques, three types of Genetic algorithm and particle of swarm algorithm. Techniques were applied on a case study of network road which contains 13 traffic lights. We used SUMO (Simulation of Urban MObility) to simulate the network. Heuristic optimization techniques themselves need to be calibrated. Calibrating them using the real problem is time consuming because simulation is computation demanding. We tried to calibrate them using a function that is assumed to have similar response surface but lighter computation demand then use the calibrated technique to optimize the traffic lights signal timing. After some comparing processes of optimization results, we discovered that one type of GA and PS at determined parameters are more suitable to produce the minimum total travel time.
Keywords: Heuristic Optimization, Traffic lights signal timing, Genetic algorithm, Particle Swarm.

Limit Theorems for Description of the Development of Epidemic in a Large Population

Larisa Afanasyeva, Elena Bashtova
Lomonosov Moscow State University, Moscow, Russia

The report is devoted to analysis of the reliability of systems with large number of elements. We suppose that failures of the elements are connected to a certain external factor. It leads to the situation when failures of the different elements depend on each other.
The expression for a generating function of the number of working elements at any time t is established. The asymptotic analysis for large systems is given in three cases.

First, the initial number n of the elements tends to infinity and the random environment does not depend on n. The Law of Large Numbers (LLN) and Central Limit Theorem (CLT) with random normalizing coefficients are proved. Second, the random environment depends on parameter n and satisfies the LLN. The LLN and CLT are obtained for the number of faulty elements. Finally, the situation when the environment stays unchanged, but the reliability of the elements and their number are increasing is considered. Specifically, it is proved that the number of faulty elements weakly converges to a doubly stochastic Poisson process with a leading function defined by the environment. These results can be useful from applications view-point.
As example the model of the development of a plague in a large population is considered. Initially in a population of n + m people, there are m infected persons. They are spreading infection across the external environment. The time Xj during which the jth infected person can spread the infection is a random. Random variable Yj defines an immune threshold of the jth person. The one gets unwell when the influence of the external environment (level of infection in the body of the person) exceeds this threshold. Seek person is always isolated, thus he/she cannot be a source of the infection for other people.
Keywords: Reliability, Limit Theorems, Random Environment.

Statistical Modeling of Companies' Perceptions on CRM Implementation

C. Agapitou[1], S. Bersimis[2], D. Georgalelos[1], N. Georgopoulos[1]
[1]Department of Business Administration, University of Piraeus, [2]Department of Statistics and Insurance Science, University of Piraeus, Greece

This paper attempts to model companies' perceptions on CRM implementation during the fiscal crisis in Greece. Specifically, explores, through a relevant survey, companies' perceptions on Customer Relationship Management (CRM) benefits and how the fiscal crisis in Greece may influence CRM implementation. All responses were appropriately analysed using multivariate statistical techniques, and showed that most companies are satisfied by the implementation of CRM strategy. In addition, it is strongly related to crucial benefits for the company, concerning customer loyalty, cost reduction and quality improvement of the services offered to customers. Moreover, companies believe that Information Technology is critical to the successful implementation of CRM. Based on the analysis a statistical model is built which represents the dynamics among variables and their relation to CRM implementation during the fiscal crisis in Greece.
Keywords: Statistical Analysis, CRM, CRM Implementation.

IMPACTS OF DIABETES AND HOMICIDE MORTALITY ON LIFE EXPECTANCY IN MEXICO

Alejandro Aguirre[1], Fortino Vela[2]
[1]El Colegio de México, [2]Universidad Autonoma Metropolitana, Mexico

Mexico presents sharp contrasts on the evolution of the level of mortality. On one hand, chronic degenerative diseases, especially those related to diabetes, are in an alarming increase. On the other, different criminal groups associated with drug trafficking have increased the number of violent deaths in some states. This has resulted in the decline of life expectancy in certain regions. For example, in regard to the death rate from diabetes mellitus, while in 2000 this figure reached 47.8 per hundred thousand for 2011 was 71.0 per hundred thousand. Meanwhile, in regard to the mortality rate from homicide, while in 2007 it amounted to 8.2 per hundred thousand in 2010 had increased to 22.9 per hundred thousand.

The paper aims to estimate the life expectancy of the Mexican population in different scenarios of evolution of these two causes of death. Mortality tables are estimated for these two specific causes of death, considering in turn the estimated years of life lost (Arriaga, 1996). The information from vital statistics, available for 2010 at national level and by state is used. The estimations may provide clues to the Mexican authorities to formulate population, health and social policies that help to increase the limits of life expectancy in our country.

Keywords: demographic analysis, mortality by cause, life expectancy.

References

1. Arriaga, Eduardo, “Commments on some indexes to measure the level and change of mortality“. Estudios Demográficos y Urbanos, 11, 5-30 (1996).
2. Aguirre, Alejandro y Fortino Vela, “Una mirada retrospectiva a la evolución de la mortalidad y sus vínculos con la política de población en la materia en México, en los últimos 50 años“. 50 años del Centro de Estudios Demográficos, Urbanos y Ambientales, El Colegio de México, in press (2014).

Virtual Waiting Time in an Unreliable Tandem Queueing Network

Amar Aissani
Laboratoire RIIMA, Department of Computer Science, USTHB, Algeria

We consider a Closed Queueing Network in tandem with arbitrary service time distributions and FIFO queues. The servers are subject to

random breakdowns and renewalls. We obtain the probability distribution of the virtual waiting time in terms of Laplace-Stieltjes transform.
Key Words: Tandem Queueing Network, Virtual Waiting Time, Unreliable Servers, Laplace-Stieltjes Transform.

THE SPATIAL PROBIT MODEL – AN APPLICATION TO THE STUDY OF BANKING CRISES AT THE END OF THE 90'S

Andrea Amaral[1], Margarida Abreu[2], Victor Mendes[3]
[1]FEP - Faculdade de Economia do Porto, [2]ISEG – Instituto Superior de Economia e Gestão and UECE, [3]CMVM - Portuguese Securities Commission and CEFAGE-UE (Center for Advanced Studies in Management and Economics), Portugal

We use a spatial Probit model to study the effect of contagion between banking systems of different countries on the probability of a systemic crisis in one county. Applied to the late 90's banking crisis in Asia we show that the phenomena of contagion is better seized using a spatial than a traditional Probit model. Unlike the latter, the spatial Probit model allows one to consider the cascade of cross and feedback effects of contagion that result from the outbreak of one initial crisis in one country or system. These contagion effects may result either from business connections between institutions of different countries or from institutional similarities between banking systems.
JEL classification: C21, C25, G01, G21.
Keywords: spatial probit, banking crises, contagion.

*The author gratefully acknowledges partial financial support from FCT, program POCTI. The views stated herein are those of the author and are not necessarily those of the CMVM.

Monitoring the Coefficient of Variation for Short Run Production

Asma Amdouni[1;2], Philippe Castagliola[1], Hassen Taleb[3], Giovanni Celano[4]
[1]Université de Nantes & IRCCyN UMR CNRS 6597, France, [2]Institut Supérieur de Gestion, Université de Tunis, Tunisie, [3]University of Gafsa, Tunisia, [4]Università di Catania, Italy

Monitoring the coefficient of variation is an effective approach to Statistical Process Control when the process mean and standard deviation are not constant but their ratio is constant. The CV has several applications. For instance, the CV is commonly used in renewal theory, queuing theory, and reliability theory. In the field of finance, it is interpreted as a measure of the risk faced by investors, by relating the

volatility of the return on an asset to the expected value of the return. It is also adopted in chemical and biological assay quality control to validate results. It can also be used in the fields of materials engineering and manufacturing where some quality characteristics related to the physical properties of products constituted by metal alloys or composite materials often have a standard-deviation which is proportional to their population mean. These properties are usually related to the way atoms of a metal diffuse into another. Tool cutting life and several properties of sintered materials are typical examples from this setting.
Until now, research has not investigated the monitoring of the coefficient of variation for short production runs. Viewed under this perspective, this presentation investigates new methods to monitor the coefficient of variation for a finite horizon production by means of one-sided Shewhart-type and Run-Rules-type charts. The truncated run length properties of these charts are derived and Tables are provided for the statistical properties of the proposed charts when the shift size is deterministic. Illustrative examples are given in order to illustrate the use of these charts on real data.

Demographic characteristics of Hypoaspis aculeifer fed on Rhizoglyphus echinopus at constant temperatures

Mohammdreza Amin, Mohammad Khanjani, Babak Zahiri
Department of Plant Protection, Faculty of Agriculture, Bu-Ali Sina University, Iran

Hypoaspis aculeifer (Canestrini) is a polyphgous mite and plays an important role in natural control of terricolous insects, mites, nematodes and fungus. Immature and adult survival and stadia as well as daily fecundity of this predator were examined in germinators at 11 constant temperatures (15, 16, 17.5, 20, 22.5, 25, 27.5, 30, 32.5, 33.5, and 35 °C), relative humidity of 60 ± 5 % and photoperiod of 16:8 hours (light:dark) feeding on bulb mite, Rhizoglyphus echinopus (Fumouze and Robin), as prey. The demographic parameters were analyzed using the age-stage, two-sex life table with bootstrap technique. The extremes of gross reproductive rate (GRR) were reached at 16 °C and 30 °C to be 5.77 and 71.48 offspring, respectively. The net reproductive rate (R0) was 0.91, 14.67, 21.87, 24.87 and 32.38 offspring at 16, 20, 22.5, 27.5 and 30°C, respectively; however, it extremely decreased to 6.72 offspring at 32.5˚C. The mean generation time (T) decreased significantly by increasing temperature from 16 °C (63.74 days) to 30 °C (13.78 days) and then increased to 16.70 days at 32.5 °C. The finite rate (λ) were calculated to be the lowest at 16 °C (0.997 day–1) and the highest at 30 °C (1.286 day–1) followed by a decline to 1.116 day–1 at 32.5 °C. The intrinsic rate of increase (r) values ranged from -0.002 day–

1 at 16 °C to 0.252 day–1 at 30 °C. This study indicated that H. aculeifer can be successfully reared on R. echinopus at 30 °C to be applied in biological control programs.
Key words: bulb mite, fecundity, life table, predatory, survival.

Dynamic Modeling and Forecasting Stability Deterministically Chaotic Processes of Functioning of the Human Body

Valery Antonov[1], Anatoly Kovalenko[2], Sergey Nosyrev[3], Almira Ivanova[4]
[1]St. Petersburg State Polytechnical University, [2]Ioffe Physical-Technical Institute of Russian Academy of Sciences, [3]Mechnikov North-Western State Medical University, [4]St. Petersburg National Research University of Information Technologies, Mechanics and Optics, Russia

Ways and means of dynamic analysis of overall stability and operational forecasting critical destabilization of the functional state of the human body based on the synergetic concept of self-organizing open nonequilibrium thermodynamic systems They are based on fractal analysis of the generation of informational entropy in stochastic components of variation rhythms of its physiological processes. Source of diagnostic information are indicators of the population density of the phase space attractor maps time series of baseline measurements biomedical signals (electrical potentials heart rate, brain activity, breathing and etc) . Along with the entropy- fractal parameters to assess the various aspects of sustainability deterministically chaotic processes used Lyapunov stability indicators characterizing the rate of divergence or convergence initially very close phase trajectories. Also consider the stability of Lagrange (trajectory remains in the closed area), and Poisson (trajectory repeatedly returns to start point area).
Clinical studies have shown high efficiency of using specially developed for this model of software for real-time processing of electrocardiograms of patients of varying severity. Objective assessment of the dynamics of the general condition of the body contributes to proactively identify critical threats to the lives of patients with pathological and age-related changes, as well as the impact of various external and internal loads
Keywords: Dynamic modeling, stability, deterministic chaos, the critical destabilization, functional status, the human body, information entropy, countries attractor, multifractal dimension, Lyapunov stability indicators, Lagrange and Poisson

Software Package for Calculating the Fractal and Cross Spectral Parameters of Cerebral Hemodynamic in a Real Time Mode

Valery Antonov[1], Artem Zagaynov[2]
[1]Saint-Petrsburg State Polytechnical University, [2]Military Space Academy Mozhaiskogo, Russia

The aim of the work is to create an automated system of calculating the cross spectral characteristics and correlation and fractal indexes of the two time series in a real time mode. The complex processes the two time series, corresponding to values of blood flow velocity (BFV) and systemic arterial pressure (SAP).
Keywords: automated system of calculating, two time series, cross spectral characteristics, fractal indexes.

Estimation of Common Correlation Coefficient Using R

Masood Anwar
Department of Mathematics, COMSATS Institute of Information Technology, Pakistan

This paper presents four different methods of estimating a common correlation coefficient from independent samples of different sizes drawn from bivariate normal populations using R. Efficiency comparison of these estimators is done through simulation. Hypothesis testing about the homogeneity of correlations is also proposed. In order to show the utility of test two real data sets are included.
Key Words: Bivariate normal population; Correlation coefficient; Estimation

Stochastic Approximation for the Resolution of an Inverse Problem with Associated Errors

Idir ARAB, Abdelnasser DAHMANI
Department of Mathematics, University of Bejaia, Algeria

Let us consider an inverse problen formulated by the following equation:
Ax=u...(1)
where A is a compact operator in a Hilbert space H. We assume that (1) has a unique solution xex satisfying Axex=uex such that uex is unknown. In the practise the second member of eq(1) is a result of measurement, inevitably errors will take place, when carrying out n experiments we obtain a sample {U1,..,Un} under the form Uk=Uex+ek (ek is the error). Since the operator A is compact, these errors will imply a large deviation in the solution, thus requiring special techniques for solving (Approximate the exact solution) (1) in order to satisfy the stability condition. Many methods were introduced as Landweber algorithm, Tikhonov regularisation, stochastic approximation (Robbins-monro algorithm) ...etc. These algorithms assure the convergence without using

inversion of the operator A, thing that is impossible in the most cases. In this paper we introduce the Robbins-Monro algorithm : Xk+1=Xk-ak(AXk-Uk) to approximate the exact solution. In this case, the induced errors (ek) can be in a large class which is the class of associated random variables that contains independent ones. We establish the convergence of the algorithm and we obtain exponential inequalities.
In numerical scheme, we exemplify by taking the operator A, the Fredholm operator and a second member of (1) for which we have the exact solution, we compare the approximate solution and the exact one to show the accuracy of the algorithm.
Key Words: Inverse problems, Stochastic approximation, Associated random variables, Exponential inequalities.

Variance of Estimation in two-phase sampling

R. Arnab, J.O.Olaomi
Department of Statistics, University of South Africa, South Africa

The variance estimation for two-phase sampling was considered by Haziza et al. (2011). In this paper few alternative estimators have been proposed. The proposed variance estimators use auxiliary information collected in the first phase effectively to improve efficiencies. Empirical investigations reveal that the proposed variance estimators are more efficient than those estimators proposed by Haziza et al. (2011). It is also shown that all the proposed variance estimators takes negative laule much less fequent than those proposed by Haziza et al. (2011).
Key Words: Two-phase sampling, variance estimation, varying probability sampling

A bootstrap approximation for the distribution of the Local Whittle estimator

Josu Arteche, Jesus Orbe
Dpto. Economía Aplicada III. Econometría y Estadística. Universidad del País Vasco/Euskal Herriko Unibertsitatea, Spain

We have proposed in the paper at hand two bootstrap strategies for the estimation of the exact distribution of the LW estimator that outperform significantly the approximation offered by the asymptotic distribution in a wide range of situations. The asymptotic properties of the Local Whittle estimator of the memory parameter d have been widely analysed and its consistency and asymptotic distribution have been obtained for values of d in (-1/2,1] interval in a wide range of situations. However, the asymptotic distribution may be a poor approximation of the exact one in several cases, as for example with small sample sizes or even with larger samples when d>0.75. In other situations the asymptotic distribution is unknown, as for example in a noninvertible context or in some nonlinear transformations of long memory processes where only consistency has been obtained. For all these cases two different bootstrap strategies are proposed. The proposals are based on resampling two different standardized periodograms in order to obtain

bootstrap replicates of the periodogram. The first one uses a Studentized periodogram and is resampled over the whole band of Fourier frequencies in a global bootstrap strategy. The second one uses a local standardization of the periodogram and is resampled using a local bootstrap strategy. A Monte Carlo shows that this strategy leads to a good approximation of the exact distribution of the Local Whittle estimator in those situations where the asymptotic distribution is not reliable. Thus, we can see that both bootstrap strategies tend to lead to coverage closer to nominal, even with narrower intervals in some cases.
Keywords: Local Whittle estimator, Bootstrap, Confidence intervals.

Language variation after a harmonization process. A multivariate analysis for the Basque language case

G. Aurrekoetxea[1], K. Fernández-Aguire[2], J. L. Ormaetxea[1], J.Rubio[2]
[1]Manlab, Department of Linguistics and Basque Studies (UPV/EHU),
[2]Department of Applied Economics III (UPV/EHU), Spain

This research presents some results on dialect levelling in the Basque language which are closely related to the age-based sociolinguistic variation in the Basque language.
Dialect levelling in different languages and dialectal situations is a well known phenomenon, but there are only a few investigations on levelling in the Basque language. This research began ten years ago (Aurrekoetxea 2004), and has guided the creation of the EDAK corpus (The Dialectal Oral Corpus of the Basque Language). This corpus records and gathers phonological, morphological, syntactic, and lexical information from different places within the territory of the Basque language. In each of these localities, data from two generations (old and young speakers) has been gathered.
Both generations have experienced different patterns of education: older people have been educated in Spanish and have never studied any subject in Basque during their school years, whereas young people have for the most part been schooled in standard Basque. Indeed, as the standard Basque variety was created in the last decades of the twentieth century, in most cases the adult generation has had no direct contact with it, while the younger generation has been immersed in it since childhood.
The data used in this study are taken from the EDAK corpus and are based on lexical (e.g. eguzki ‘sun’, ilargi ‘moon’, euskara ‘the Basque language’), morphological, and syntactic parameters. These data show the levelling phenomenon that is taking place in the Basque language today.
Our aim is to apply quantitative methods to these data, taken from the above-mentioned sociolinguistic survey. We use multivariate statistical procedures to show the direction of the levelling process which is currently taking place in the Basque language.
Keywords: Dialectology, Multivariate Analysis, Factorial Methods, Clustering
References

1. Auer, P. & F. Hinskens and P. Kerswill (eds.), Dialect Change. The Convergence and Divergence of Dialects in European Languages. Cambridge: Cambridge University Press (2005)
2. Aurrekoetxea, G. & Iglesias, A. “Technology for prosodic variation”, in Aurrekoetxea, G. & J. L. Ormaetxea (eds.): Tools for Linguistic Variation, Bilbao: UPV/EHU, 207-229 (2010)
3. Lebart, L., Salem, A., and Berry, Exploring Textual Data. Dordrecht: Kluwer Academic Publishers (1998)

Mathematical analysis of the role of population heterogeneity in the evolution of human mortality dynamics

Demetris Avraam[1], Joao Pedro de Magalhaes[2], Séverine Arnold-Gaille[3], Bakhtier Vasiev[1]

[1]Department of Mathematical Sciences, University of Liverpool, UK, [2]Integrative Genomics of Ageing Group, Institute of Integrative Biology, University of Liverpool, UK, [3]Department of Actuarial Science, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Switzerland

Many features of biological populations can be described in terms of their heterogeneity by taking into account the variation among individuals and cohorts in the population. In demography, the heterogeneity of populations can explain various features of age-dependent demographic observations including those related to mortality dynamics. Mortality dynamics is underlined by the Gompertz law stating that the mortality rate increases exponentially with age. However deviations from the exponential increase are commonly observed at early- and late-life intervals. There have been developed a few different models (i.e Heligman-Pollard model) designed to describe these deviations. These models postulate that a few different processes take place in the population and affect its mortality dynamics. In this study we design a model based on an assumption that the mortality dynamics is indeed underlined by a Gompertz law and deviations from this law are due to the heterogeneity of human populations. The model manages to replicate the entire pattern of mortality and to explain the deviations from the exponential growth. We use fitting techniques to match the model output to observed data on age-dependent mortality rates for Swedish population. We show that the model of heterogeneous population composed by four subpopulations, each following a Gompertz law, fits the data very well. We also transfer the idea of heterogeneity to probability density and survival functions to fit deaths and survivors’ numbers of the Swedish population. Finally, we analyse the time-evolution of the model parameters that best fits the Swedish mortality rates for consecutive years within a century. The analysis of the evolution of model parameters confirms the compensation law of mortality for each subpopulation and also indicates a process of homogenization of the population.

Keywords: Gompertz law of mortality, Heterogeneity, Mathematical model, Model fitting, Compensation law of mortality

Bayesian reconstruction of subsurface conductivity from geosounding data

Robert G Aykroyd[1], Hugo Hidalgo-Silva[2], Enrique Gómez-Trevino[3]
[1]Department of Statistics, University of Leeds, UK, [2]Departamento de Ciencias de la Computación, Centro de Investigación Cientfica y Educación Superior de Ensenada, México, [3]Departamento de Geofsica Aplicada, Centro de Investigación Cientfica y Educación Superior de Ensenada, México

The investigation of the earth's internal structure is of great interest in areas as diverse as civil engineering, water management and oil exploration. A common technique looks to estimate the electrical conductivity of the subsurface, a proxy for the internal structure, from a grid of surface voltage measurements - an example of an ill-posed inverse problem. This paper uses a Bayesian modeling approach to compare different prior model descriptions using Markov chain Monte Carlo estimation. A range of simulation examples are considered.
Keywords: Bayesian modelling, Inverse problems, MCMC, Posterior estimation, Regularization.

A data mining method to measure accuracy in child speech

Elena Babatsouli
University of Crete, Greece

It is well known that consonants as opposed to vowels are the main cause for erroneous speech during child phonological development which normally completes around school age. Accuracy of an individual consonant's production is measured per word and then arithmetically averaged over all words since it has been established in the literature that the typology of a word (number of syllables, consonant clusters, etc.) usually affects the accuracy of consonant production. This measurement is called proportion of consonant correct (PCC). In practice, however, it is a daunting task to collect all the tokens of each word type in a child's speech and then measure consonant production accuracy. The purpose of the present paper is to propose a data mining method in order to make life easier for data collectors and speech evaluators by reducing the amount of speech data to be collected and measured. In the proposed method, each word's tokens are considered but not measured. If the consonant under examination is produced by the child at least once correctly in a specific word, then the consonant is given 1 point for this word type, otherwise it is given 0 points. This way, measuring the consonant's accuracy in all word tokens is avoided. As a result, each consonant's accuracy cumulatively for all word types is measured by the proportion of correct word types to all word types, called here proportion of types correct (PTC). At different stages during phonological development, it is possible that the two measurements PTC and PCC coincide for some consonants. This is certainly true for consonants that are either never produced correctly or are always produced correctly. What happens for the remaining cases? If PTC is measured can PCC be predicted by a model and if so how accurately?

The proposed method is applied to the data collected by the author from her child's English speech at the age of two years and seven months. The speech data considered comprises 317 word types with 1,516 tokens and 24 different consonants with 2,944 tokens. The resulting PTC has a mean of 0.457 and a variance of 0.124. If PTC is modeled as equal to PCC, the coefficient of determination (R2) of PTC is 0.871, for this stage of the child's phonological development. On the other hand, a linear model of best fit between PTC and PCC yields R2=0.933. A 6th degree polynomial approximating PTC as a function of PCC yields R2=0.978. Further, a simple model found for the relationship between PTC and PCC is PTC=PCC0.6 yielding R2=0.972 for this approximation. In a similar manner, model norms can be established by considering data from several children at different stages of speech development, so that from then on a child's PCC will be predicted for each consonant based on the measured PTC. It is concluded that the proposed data mining method can satisfactorily measure consonant accuracy in child speech, substantially reducing the amount of speech transcriptions to be performed.

Comparison of BMA and EMOS statistical calibration methods for ensemble weather prediction*

Sándor Baran[1], András Horányi[2], Dóra Nemoda[1]
[1]Faculty of Informatics, University of Debrecen, [2]Hungarian Meteorological Service, Budapest, Hungary

The evolution of the weather can be described by deterministic numerical weather forecasting models. Multiple runs of these models with different initial conditions result in forecast ensembles which are used for estimating the distribution of future atmospheric variables. However, these ensembles are usually under-dispersive and uncalibrated, so post-processing is required.
In the present work we compare different versions of Bayesian Model Averaging (BMA) and Ensemble Model Output Statistics (EMOS) post-processing methods in order to calibrate 2m temperature and 10m wind speed ensemble forecasts of the ALADIN Limited Area Model Ensemble Prediction System of the Hungarian Meteorological Service (Horányi et al., 2011). Both approaches provide estimates of the probability density functions of the predictable weather quantities, where model parameters are estimated using forecast ensembles and validating observations from a given training period (Sloughter et al., 2010; Thorarinsdottir and Gneiting, 2010).
First we determine the optimal lengths of the training periods for both weather quantities by comparing the goodness of fit of models with training periods ranging from 10 to 60 days (Baran et al., 2013). Then we check the fit of the optimal models and with the help of appropriate scoring rules we show that compared to the raw ensemble both post-processing methods improve the calibration of probabilistic and accuracy of point forecasts and that the best BMA method slightly outperforms the EMOS technique.
Keywords: Bayesian model averaging, ensemble model output statistics, ensemble calibration, scoring rules.

References
1. Baran, S., Horányi, A., Nemoda, D., Statistical post-processing of probabilistic wind speed forecasting in Hungary. Meteorol. Z. 22, 273-282 (2013)
2. Horányi, A., Mile, M., Szűcs, M., Latest developments around the ALADIN operational short-range ensemble prediction system in Hungary. Tellus A 63, 642-651 (2011)
3. Sloughter, J. M., Gneiting, T., Raftery, A. E., Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. J. Amer. Stat. Assoc. 105, 25-37 (2010)
4. Thorarinsdottir, T. L., Gneiting, T., Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression. J. Roy. Statist. Soc. Ser. A 173, 371-388. (2010).

*This research has been supported by the Hungarian Scientific Research Fund under Grant No. OTKA NK101680/2012 and by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

Optimal designs for parameter estimation and prediction of shifted Ornstein-Uhlenbeck sheets*

Sándor Baran[1], Kinga Sikolya[1], Milan Stehlík[2]
[1]Faculty of Informatics, University of Debrecen, Hungary, [2]Institut für Angewandte Statistik, Johannes Kepler University in Linz, Austria

Measurement on the regular grid can be of interest for many important applications, e.g. measurement in chemistry. Our aims to contribute here for the case of methane modelling in troposphere, lowest part of atmosphere. We concentrate a realistic parametric model of covariance and provide theoretical and numerical results on optimal designs. We derive exact optimal designs for estimation of parameters of a shifted Ornstein-Uhlenbeck sheet (Baran et al., 2013; Baran and Stehlík, 2013) and of an Arrhenius model with Ornstein-Uhlenbeck error (Rodríguez-Díaz et al., 2012) observed on a regular grid. Moreover we study the optimal design problem for prediction of a shifted Ornstein-Uhlenbeck sheet with respect to integrated mean square prediction error and entropy criterion. Finally, we illustrate applicability of obtained designs for increasing/measuring the efficiency of the engineering designs for estimation of methane rate in various temperature ranges and under different correlation parameters. We show that in most situations the obtained designs have higher efficiency.
Keywords: Arrhenius model, correlated observations, optimal design of experiments, Ornstein-Uhlenbeck sheet, integrated mean square prediction error, entropy, Fisher information.
References
1. S. Baran, K. Sikolya, M. Stehlík, On the optimal designs for prediction of Ornstein-Uhlenbeck sheets, Statist. Probab. Lett. 83 (2013) 1580--1587
2. S. Baran, M. Stehlík, Optimal designs for parameters of shifted Ornstein-Uhlenbeck sheets measured on monotonic sets, TEST (2013), submitted.

3. J.M. Rodríguez-Díaz, M.T. Santos-Martín, H. Waldl, M. Stehlík, Filling and D-optimal designs for the correlated Generalized Exponential models, Chemometr. Intell. Lab. Syst. 114 (2012) 10--18

* This research has been supported by the Hungarian –Austrian intergovernmental S&T cooperation program TÉT_10-1-2011-0712 and partially supported the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

Monitoring Multivariate Dispersion and Interpreting Out-of-Control Signals

Georgios Bartzis
Department of Statistics and Insurance Science, University of Piraeus, Greece

Multivariate Statistical Process Control is used for monitoring and controlling a process, when the quality of the process is related p>1 correlated variables. Furthermore, a rather interesting problem regarding multivariate statistical process control is the interpretation of an out-of control signal which means the identification of the responsible variable(s) which led the process to be out-of-control. To overpass the identification problem, the researchers have proposed interesting techniques over the years. In this paper, we initially present an extensive literature review for monitoring the process dispersion and finally we describe different methods for interpreting the out-of-control signal of a multivariate control chart for process dispersion.
Keywords: Interpretation, Multivariate Control Charts, Dispersion, Outbreak Detection

Assessment of Classical and Bayesian approach for Estimation of Structural changes in Panel Data

Ishita Basak, Ashis Kumar Chakraborty
SQC & OR Unit, Indian Statistical Institute, India

In various applications, the data obtained during an extensive time period has to be investigated representing that probable statistical model may alter once or several times during the period of surveillance. The alteration in statistical model signifies a change and the point of alteration occurrence is the change point. In a sequence of random variables $X_1, X_2, \ldots\ldots X_n$ if $X_1, X_2, \ldots\ldots X_\eta$ $(\eta < n)$ follows a common distribution F and $X_{\eta+1}, X_{\eta+2}, \ldots\ldots X_n$ have distribution G with F ≠ G, then, the index 'η' is called the change point. Statistical exploration of change-point problems depend on the type of data to be examined. Time series data is usually of two types: one-dimensional and multi-dimensional or panel data. A 2-D data quite often, may be a panel data gathered at several time points on numeral subjects. Introduction of a shared involvement for all the themes (subjects) may lead to single or

multiple change points occasionally called structural changes, in each row of panel data.
The current research focuses on different general modelling approaches to determine the presence or absence of multiple change points in each row of the 2-D data congregated at numerous time points on different subjects. Two approaches, Classical and Bayesian were implemented to estimate the change points in the per capita income change of 50 US states observed from 1948 to 2013. The Classical approach was applied following Random effects model with three terms in consideration *viz.*, common term for all subjects, subject specific error term and individual error term. Estimation of change points were done by least square theory. The Bayesian approach was employed concerning three assumptions; (a) subjects follow normal distribution having a change in parameters after the change points, (b) a correlation exists among the parameters before and after the change point for each subject, (c) change points of different series follow a common distribution. In this approach, estimation was performed by Markov Chain Monte Carlo (MCMC) method. A comparison amid the two estimates was executed by determining the standard error.
In conclusion, two change points were observed in many of the states, generally, in 1984 and 1988. Some of the states exhibited no evidence of structural changes implying diminished effect of Great Moderation. The Bayesian approach displayed better estimate over the Classical one.
Keywords: Panel data, Change point, Classical approach, Least square theory, Bayesian approach, Markov Chain Monte Carlo method.

Mathematical Models for Calculation of Relative Growth Rate using Seedling Length

Behnam Behtari, Kazem Ghassemi Golezani
Dep. of Crop Ecology, Faculty of Agriculture, University of Tabriz, Iran

Seedling emergence is one of the most important biological processes determining crop yield. Relative growth rate have estimated by the production of dry matter and it can be used to compare the species efficiency or it can be applied to determine temperature effects in certain conditions. The objective of this research was to determine the relative growth rate, using seedling length and field emergence data of three crop species including green bean, sunflower and maize. To do this, an experiment based on randomized complete block design was conducted at the Research Farm of Faculty of Agriculture and Natural Resource College of Ahar, University of Tabriz. Based on observation data, the two methods were used to calculate of relative growth rate (RGR). In first method, RGR was calculated by Hunt equation. In the second method, we used of fitting the Logistic and Gomprtz nonlinear regression

equations that had s! hown the best fit compared with others. In general, 24 models were fitted and the multimodel inference was used to calculate the models probability and averaging. Analysis of variance to compare of the three methods of RGR among plants, shown significant differences, however, all models that make from actual data of greenhouse experiment was used to validation. The results of validation showed that there is not significant difference between RGR of Hant and Logistic models; indicate the suitability of these models to calculate RGR by seedling length

Key Words: Logistic model, Gomprtz model, Hunt model, Length of seedling, Relative growth rate.

Path and ridge regression analysis of seed yield and seed yield components of soybean under different irrigation regimes

Behnam Behtari
Dep. of Crop Ecology, Faculty of Agriculture, University of Tabriz, Iran

Seed yield, a quantitative character, is largely influenced by the environment, and thus has a low heritability. Therefore, the response to direct selection for seed yield may be unpredictable unless environmental variation is well controlled. The objective of this study was to examine the mathematical relationships between seed yield and its components by using a path analysis and ridge regression modeling approach to forecast the seed yield in seed production. To do this, a split-plot experimental based on randomized complete block design with three replication was conducted in 2004. Irrigation treatments were assigned to main plots, and two soybean cultivars were allocated to the subplots. Irrigation treatments I1, I2, I3 and I4 were defined based on the cumulative evaporation of 60±3, 80±3, 100±3 and 120±3 mm, from pan (class A), respectively. The seed yield components considered in this study, were num! ber of pods per plant (x1), number of seeds per pod (x2), number of seeds in plant (x3), pod-bearing nods in plant (x4) and seed weight (x5). Pearson correlation coefficients and path analysis of components x1 through x5 to Y showed that the strongest indirect effect on Y was x1 via x3 (the coefficients is 0.29), x1 via x4 (the coefficients is 0.24) and x3 via x1 (the coefficients is 0.24). All of the ridge coefficients were positive except x2 that in two levels of I1 and I4 irrigation treatments was negative. This result showed that number of seeds per pod trait in the soybean crop is very sensitive to the amount of water so that with the maximum and minimum amounts of water, its value was related to yield inversely. This study developed an original exponential model for estimating yield from the values of yield components under influenced different irrigation regimes. The model was statistically reliable. In the present trial, the genetic controls were more general than

the environmental controls for x1 to x5. Therefore, we tentatively propose that x1, x3 and x4 were orderly more genetic and less environmental control than x2 and x5.

Key Words: Exponential model, Multicollinearity, Path analyses, Ridge regression

Stochastic processes associated with general systems of nonlinear parabolic equations

Yana Belopolskaya
St.Petersburg State University for Architecture and Civil Engineering, Russia

Systems of nonlinear parabolic equations arise as mathematical models of various phenomena in various problems of chemical and biological nature as well as theory conservation laws, control theory, financial mathematics and others. We derive some probabilistic approaches to construct probabilistic representations for classical, generalized and viscosity solutions of the Cauchy problem to these systems. Actually, we consider two sets of systems of nonlinear parabolic equations, namely weakly coupled systems, which are diagonal in higher order terms and fully coupled systems with nondiagonal entrance of higher order terms.

To solve the Cauchy problem for a system of the first type we have to construct diffusion processes and their multiplicative operator functionals. We construct them as solutions of the corresponding stochastic equations with coefficients depending on distributions of their solutions. Finally, the solutions of these SDEs are used to derive stochastic representations for classical solutions of the original Cauchy problem. To construct generalized solutions of PDE systems from this class we need additionally some results of the theory of stochastic flows while to deal with viscosity solutions we turn to the technique of backward SDEs

For systems of the second type, we derive stochastic representations for both classical and generalized solutions. To deal the Cauchy problem solutions in this case we need even more sophisticated approach including both stochastic flow theory and BSDE theory.

Probabilistic representations of the Cauchy problem solution for systems of parabolic equations together with Monte Carlo technique allows to develop rather effective numerical algorithms to construct the required numerical solutions.

This work is partly supported by grant RFBR 12-01-00457 and NIR Minobrnauki 1.370.2011

References

1. Belopolskaya, Ya. and Woyczynski W., Generalized solution of the Cauchy problem for systems of nonlinear parabolic equations and diffusion processes, Stochastics and dynamics, v 11 N 1, 1-31 (2012).
1.Ya.Belopolskaya, Probabilistic counterparts of nonlinear parabolic PDE systems Modern Stochastics and Applications - Springer Optimization and Its Applications 90 (2014)

Monitoring Multivariate Non-Industrial Processes

Sotiris Bersimis, Athanasios Sachlas
Department of Statistics and Insurance Science, University of Piraeus, Greece

Woodall and Montgomery (1999) stated that MSPC is one of the most rapidly developing areas of SPC. Based on this fact, Bersimis et al. (2007) gave an extensive literature review of MSPC. Open problems in the field of MSPC are related to non-parametric procedures, interpretation of out-of-control signals, autocorrelation, etc.
Nowadays, the use of MSPC techniques is efficiently generalized beyond product quality and the monitoring of industrial processes, in order to be used in many other fields (e.g. Public-Health, Environmental, Financial monitoring, etc). Non-Industrial processes are usually non-normal and autocorrelated. In this paper, we focus on presenting a technique for monitoring non-industrial processes, in which more than one correlated variables are of interest. The technique is exhibited using real data.
Keywords: Multivariate control charts, Autocorrelated process, Non-industrial processes

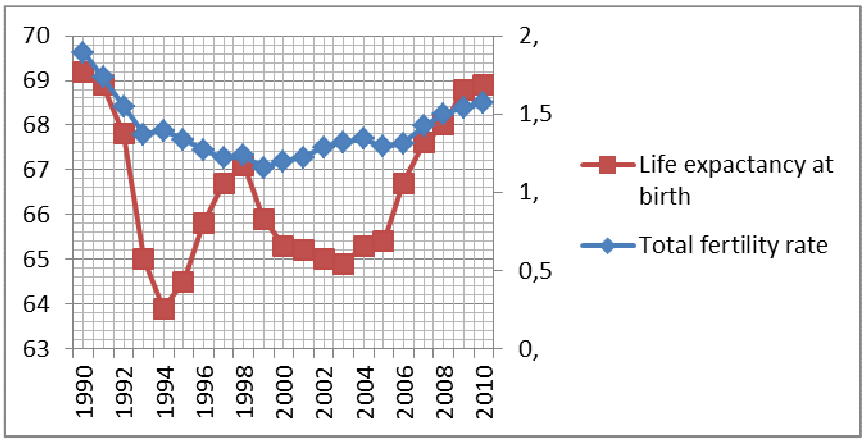
Demographic Crisis in Russia In the Frame of Demographic Transition Model

Vadim Bezverbny
Socio-Political Research Institute of Russian Academy of Sciences, Russia

The major reason of the demographic crisis in Russia remains unclear and highly debatable. A number of Russian demographers view the dramatic decline of birth rate and high percentage of nuclear families in the population of Russia as a natural result of a demographic transition (Vishnevsky, Zakharov). According to such scientists, Russia has followed the course of western countries where postindustrial economy produced a negative effect on birth rates. Other scientists (Riazantsev,

Rybakovsky) believe that demographic crisis in Russia is caused by the results of social and economic reforms in the early 1990.
In this case a period of intensive decline in birth rates which happened since 1990 to 1993 wasn't long enough for the evolutionary saltation from traditional to modern stage of demographic transition (Figure 1).

Figure1. Dynamics of the total fertility rate and life expectancy in Russia (1990-2010)



Calculated on the basis of the Federal State Statistics Service of the Russian Federation

Many researches forget about the fact that apart from the dramatic decline of birth rates the modern stage of demographic transition is also accompanied by a sharp decline of death rates and prolongation of life expectancy while in Russia we can see a totally diverse dynamics. Therefore, such hypothesis on consistency of catastrophic depopulation of Russia has good reasons in its base.

Aims:

1) To show the dynamics of changes in replacement of the Russian population in the period of 1980-1990;
2) To investigate the origins of the demographic crisis in Russia;
3) To explore level of effect of social and economic distress on the parameters of demographic evolution of Russia;
4) To esteem the demographic results of economic reforms of the early 1990;
5) To examine the demographic dynamics in Russia in the frame of demographic transition model.

Data: The data will come from the official statistics of the Russian Federal State Statistics Service.

Methods: The methodology is based on the methods of the statistical and graphical analyses. I will further use the econometrical analysis to

measure direct and indirect demographic losses of Russian population in 1990-2000.
Conclusions: There is a high possibility that chaotic economic reforms and unprecedented impoverishment of the population, disintegration of the social safety net, dramatic deterioration of living conditions and nutrition quality, loss of confidence in the future, stratification of society and loss of moral and spiritual values were the main cause of the demographic crisis in Russia.

Cogarch(1,1) model: applications and performance in analysing real data

Enrico Bibbona, Ilia Negri
Department of Engineering, University of Bergamo, Italy

The COGARCH (COntinuous Generalized Auto-Regressive Conditional Heteroschedastic) model can be considered as a continuous version of the well known GARCH discrete time model.
They are driven by general Lèvy processes and the resulting volatility process satisfies a stochastic differential equation. The main difference between COGARCH models and other stochastic volatility models is that there is only one source of randomness (the Lèvy process) and all the stylized feature are captured by the dependance structure of the model as in the GARCH models.
The prediction-based estimating functions (PBEF's) method is applied to draw statistical inference about the COGARCH(1,1) model from discrete observed data. PBEF's can be seen as a generalization of martingale estimation functions. They are based on linear predictors, have some of the most attractive properties of the martingale estimating functions, moreover an optimal prediction-based estimating function can be found.
A package for simulation and estimation of COGARCH(1,1) is presented. The aims of this work is to investigate how COGARCH(1,1) models perform on real and simulated data compared with other models both discrete than continue. We will consider different estimations method for COCARCH(1,1) models. Beyond PBEF's, a pseudo-maximum likelihood (PML) and a method of moments (MM) will be applied.
Key Words: COGARCH(1,1) models, Prediction based estimation functions, Simulations

The Cauchy Problem for BBGKY Hierarchy of Quantum Kinetic Equations with Yukawa Potential

Nikolay N.Bogolubov[1], Mukhayo Yu.Rasulova[2], Umar O.Avazov[2]
[1]V.A.Steklov Institute of Mathematics of the Russian Academy of Sciences, RUSSIA, [2]Institute of Nuclear Physics Academy of Sciences Republic of Uzbekistan, UZBEKISTAN

The existence of a unique solution, in term of initial data of the Bogolubov-Born-Green-Kirkwood-Yvon's (BBGKY) hierarchy of quantum kinetic equations with Yukawa potential is proved. This is based on non-relativistic quantum mechanics and application of semigroup theory method.
Keywords: BBGKY hierarchy, Yukawa potential, Semigroup theory

Wald Test and Distance-Based Generalized Linear Models. Actuarial Application

Eva Boj, Teresa Costa, Josep Fortiana, Anna Esteve
Departament de Matematica Economica, Financera i Actuarial, Universitat de Barcelona, Spain

The Distance-based generalized linear model (DB-GLM, Boj et al. 2012) extends the ordinary GLM, allowing information on predictors to be entered as interdistances between observation pairs instead of as individual coordinates. In turn, these interdistances may have been computed from arbitrary, non-numerical observed predictors. The estimation process of a DB-GLM is schematically as follows: a Euclidean configuration is obtained by a metric multidimensional scaling-like procedure, then the linear predictor of the underlying GLM is a linear combination of the resulting Euclidean coordinates, latent variables in the model. Therefore influence coefficients of the original observed predictors cannot be computed as in the ordinary GLM.
In a previous work presented at the ASMDA2013 conference (Costa et al. 2013) we proposed a definition of local influence coefficients for the DB-GLM model depending on the nature of risk factors (numerical or categorical/binary).
Now we study how to adapt the Wald test of predictors significance to the DB-GLM environment. To this end, firstly we use the definition of influence coefficients given in Costa et al. (2013) to estimate coefficients and then we adapt the pairs bootstrap methodology to estimate the coefficients distribution. In this way we are able to estimate the coefficients estimates standard error and the contrasts p-values. The Wald statistic is adapted to test a null hypothesis that is not true in the bootstrap data generation process.

We illustrate the use of this Wald test for selecting risk factors in rate-making. Calculations are implemented with the dbstats R package (Boj et al. 2012).
This research is part of the project: Semiparametric and distance-based methodologies with applications in bioinformatics, finance and risk management (grant MTM2010-17323).
Key Words: Distance-based generalized linear model; influence coefficients; Wald test; actuarial science.
References
Boj, E., Delicado, P., Fortiana, J., Esteve, A. and A. Caballé (2012). Local distance-based generalized linear models using the dbstats package for R. Documentos de Trabajo de la Xarxa de Referència en Economia Aplicada (XREAP), XREAP2012-11. (Submitted, in second revision, to the Journal of Statistical Software)
Costa, T., Boj, E., Fortiana, J., Esteve, A. (2013). Assessing the importance of risk factors in distance-based generalized linear models. 15th Applied Stochastic Models and Data Analysis International Conference (ASMDA2013). Barcelona (ESPAÑA), June 25-28, 2013.

On the asymptotic behaviour of a dynamical version of the Neyman contagious point process

K. Borovkov
The University of Melbourne, Australia

We consider a finite point process in evolving in discrete time as follows. Starting with an arbitrary initial configuration of finitely many points with non-negative weights, at each step a point is chosen at random from the process according to the distribution with probabilities proportional to the points' weights. Then a random number of new points is added to the process, each displaced from the location of the chosen point by a random vector and assigned a random weight. Under broad conditions on the random sequences of the numbers of newly added points, their weights and displacement vectors, we obtain the asymptotic behaviour of the point last added to the process and also that of the scaled mean measure of the point process as the number of steps goes to infinity.
Key Words: Neyman contagious point process, preferential attachment, random environments, limit theorems

Programming the Convexity Ratio and Applications

James R. Bozeman, Kelly Butler, Matthew Davey, Kayla Flynn
Mathematics and Computer Science, Lyndon State College, USA

In this talk we exhibit a computer program which finds (or approximates) a largest convex set (the endogon) in the interior of a simple polygonal

region P in the plane. The program then calculates (or approximates) the area of the endogon and finds the convexity ratio CR(P) = (area of endogon)/(area of exogon) where the exogon is the convex hull of P. This number is then used to decide whether or not current (or proposed) legislative districts in the United States may have been politically gerrymandered.

Key Words: convexity ratio, endogon, convex hull, gerrymandering, legislative districts

Optimization of Multi-component Inventory and Insurance Systems

Ekaterina V. Bulinskaya
Lomonosov Moscow State University, Russia

In this paper we present and analyze some applied probability systems having a complex structure. Thus, in order to construct a realistic model of insurance company functioning one has to take into account many factors. It is well known that the primary aim of any insurance company is satisfaction of policyholders claims. That means, for more than a century the ruin probability was considered as objective function evaluating the performance quality. In other words, the reliability approach was and still is popular in actuarial mathematics. On the other hand, being a joint-stock company, its secondary but very important task is protection of the shareholders interests by payment of dividends. So, there exists the cost approach initiated by the De Finetti seminal paper [1]. Moreover, to minimize the risk and avoid the ruin insurer can choose various methods. Namely, it is possible to buy reinsurance, make investments in risky and non-risky assets or sell the company assets at the stock market, and take bank loans, see, e.g., Bulinskaya [2]. So, we deal with a multi-component system of complex structure. The asymptotic analysis of such systems arising also in inventory theory enables us to establish optimal and asymptotically optimal policies as the planning horizon tends to infinity. We study also the systems stability with respect to small fluctuations of parameters and perturbations of underlying processes. This is useful for systems optimization under assumption of incomplete information.

Keywords: Stochastic modeling, Complex Systems, Optimal and asymptotically optimal policies, Stability, Incomplete information

References

1. De Finetti, B., Su un'impostazione alternativa della teoria collettiva del rischio. In: Transactions of the XVth International Congress of Actuaries, 2, 433-- 443 (1957).

2. Bulinskaya, E., Asymptotic Analysis of Insurance Models with Bank Loans. In: New Perspectives on Stochastic Modeling and Data Analysis. Ch. Skiadas, et al. Eds (to appear)

The continuum of antenatal care throughout pregnancy and its effect on neonatal survival in the Philippines: Testing the WHO recommendation

Filipinas Bundoc
University of the Philippines School of Economics, Philippines

Progress reports on the Millennium Development Goals show that in order to meet the target on child mortality focus must be given on reducing neonatal mortality. In the attempt to understand the issue on neonatal death, this paper looks at the importance of attaining appropriate level of antenatal care coverage throughout pregnancy based on the recommendations of World Health Organization in creating better pregnancy outcomes.

Medical literature shows that certain risks are specific to a particular period in the pregnancy. In the first trimester, the risk of spontaneous abortion due to chromosomal anomalies is high. While hypertensive disorder related pregnancy problems like pre-ecclamsia are diagnosed during the third trimester. This paper then does not just look at the timing and frequency of check-ups per se but looks at the continuum of care throughout pregnancy based on the recommendations of the World Health Organization.

Using 2008 Philippine demographic health survey dataset, results indicate that after controlling for the maternal and fetal characteristics, children born of women who completed the recommended check-ups and delivered in a facility are likely to survive neonatal period. Those of women who had late initiation of care but fulfilled all other check-ups thereafter had better neonatal outcomes compared to those who did not comply at all. These results suggest the need to encourage women to obtain appropriate level of care and deliver in a facility. Concerted effort must be made to convince expectant mothers to complete the necessary check-ups.

Key Words: child health, antenatal care, neonatal survival

Indicator based safety assessment of multimodal traffic incidents

Geltmar von Buxhoeveden, Eckehard Schnieder
Institute for Traffic Safety and Automation Engineering Technische Universitat Braunschweig, Germany

Traffic safety assessment serves as a basis for evaluation, trend identification and also for supervision of operators and their investment allocation. In order to compare traffic mode related safety quantities the Swiss Federal Office of Transport (FOT) uses a system of safety indicators to categorize public transport incidents. To understand,

analyse and communicate the risks of public transport usage in Switzerland a software tool was developed. Indicators are evaluated on a country and company level. Confidence intervals (CI) are used to compare computed safety targets from the past to the safety performance of recent years. Standardisation of results can be performed to better compare companies and groups of companies. Ranking of company safety performance is used to help safety auditors perform their job. Challenges of a three year lasting project from data collection and data cleaning up to statistical analyses, leading to a software tool that is in productive use at the FOT are described. The paper comprises a collection of recent statistical methods in practical use at a governmental institution.

Keywords: railway, traffic, transportation, accident, Switzerland, data analysis, R, shiny package.

Parameter dentification for a stochastic growth model with extinction

Fabien Campillo[1], Marc Joannides[2], Irène Larramendy-Valverde[2]
[1]Modemic / Inria, [2]I3M / Université Montpellier 2, France

We consider a stochastic logistic growth model given by a stochastic differential equation in which the birth rate and the death rate appear in both drift and diffusion terms. This model arises naturally as a diffusion approximation of a continuous time Markov chain featuring a logistic death mechanism. The information on the demographic noise contained in the diffusion term allows inference on both rates separately. We study the complete Fokker-Planck equation governing the time evolution of the process and propose numerical approximations, paying a particular attention to the extinction question. We illustrate these methods using simulated data with or without extinction.

Key Words: population dynamics, Stochastic differential equation, Fokker-Planck equation, extinction, Monte Carlo, Identifiability, logistic model

The Portuguese Stable Equivalent Population - a study of the statistical trends in the last four decades

Pedro Campos
LIAAD /INESC TEC and Faculty of Economics of Porto, Portugal

Stable population theory provides a broadly useful mathematical framework for studying population's age structure and growth that are mainly determined by rates of fertility, mortality and migration. This paper focuses on the application of the Stable model theory to the Portuguese population between 1981 and 2011 and investigates the determinants of

the evolution of the Portuguese Stable Equivalent Population. An Agent-Based model is used to simulate alternative values for the Stable model, and, consequently, the properties of ergodicity and model stability are analyzed.
Keywords: Stable Population Model, Agent-Based Models, Ergodicity

Perturbation Methods for Pricing European Options in a Model with Two Stochastic Volatilities

Betuel Canhanga[1], Sergei Silvestrov[2], Anatoliy Malyarenko[2], Ying Ni[2]
[1]Eduardo Mondlane University, Mozambique, [2]Mälardalen University, Sweden

Financial models have to reflect the characteristics of markets in which they are developed to be able to predict the future behavior of a financial system. The nature of most trading environments is characterized by uncertainty that is expressed in mathematical models in terms of volatilities. In contrast to the classical Black–Scholes model with constant volatility, our model of mean-reversion type includes two stochastic volatilities, a fast-changing and a slow-changing; this can be interpreted as the effects of weekends and the effects of seasons of the year (summer and winter) on the asset price.
We perform explicitly the transition from the historical probability measure to the risk-neutral one by introducing the two market prices of risk and applying Girsanov Theorem. To solve the boundary value problem for the partial differential equation that corresponds to the case of a European option, we perform both regular and singular multi-scale expansion in fractional powers of the speed of mean-reversion factors and construct an approximate solution given by the classical two dimensional Black–Scholes model plus some terms that expand the results obtained by Black and Scholes. Concrete examples are presented.
Keywords: Financial market, Mean reversion volatility, Risk-neutral measure, Partial differential equation, Regular perturbation, Singular perturbation, European option

Nonparametric Design of the Combined X-S Control Chart for Retrospective Analysis

Giovanna Capizzi, Guido Masarotto
Department of Statistical Sciences, University of Padua, Italy

Statistical process control is usually implemented in two phases. In Phase I, also called retrospective stage, a fixed-size sample of time-ordered data is collected and analyzed for understanding process variability, checking stability and defining the In-Control (IC) state. Then, once the IC state is established, an IC reference sample is used to set up a control chart for the Phase II prospective monitoring. Shewhart control charts based on subgroup means, ranges and standard deviations are often recommended for Phase I analysis since they are easy to understand and operate. Traditionally, the control limits of these schemes have been computed so that, exactly or approximately, a prescribed probability of giving a false alarm is achieved for normally distributed observations. However, this approach leads to an excessive number of false alarms if data are not normal. Since it is difficult to check the normality assumption if stability has not yet been established, we will investigate an alternative approach for setting up a combined X-S which can be used for detecting location and scale shifts. The suggested control limits guarantee a desired probability of false alarm without requiring any prior information on the process distribution. Simulation will be used to compare the resulting chart with other parametric and non parametric retrospective Shewhart-type schemes.
Keywords: Change Points, Distribution-Free Methods, Phase I Analysis, Statistical Process Control.

The influence of classes' entropy and overlap on Random Forests' performance

Margarida G. M. S. Cardoso[1], Maria José Amorim[2]
[1]UNIDE and Dep. of Quantitative Methods, ISCTE Business School, Lisbon University Institute, [2]Department of Mathematics, Lisbon Superior Engineering Institute, Portugal

In order to evaluate the impact of classes' overlap and entropy on the performance of Random Forests, we conduct some experiments based on synthetic data – 360 data sets are generated. We set up the scenarios for our experiments by considering different classification problems with 2, 3 and 4 classes and diverse degrees of classes' entropy and overlap. According to the obtained results, the average performance of random forests significantly decreases with the increase of the degree of classes' overlap and this impact clearly surpasses the

impact of the classes entropy. Statistical analysis conducted yield additional insights referring to diverse measures of classification performance.

On Matrix-Exponential Distributions In Risk Theory

Carleo Alessandra, De Medici Giovanna, Manca Raimondo, Pietroluongo Mariafortuna
Università di Roma "La Sapienza", Italy,

In this paper, a particular class of matrix-exponential distributions is described, also with respect to its use in risk theory, namely phase-type distributions. Phase-type distributions have the important advantage of being suitable for approximating most of other distributions as well as being mathematically tractable.
After a review on phase-type distributions and their properties, a possible use in risk theory is illustrated. Modelling both inter-arrival claim times and individual claim sizes with this class of distributions an explicit formula for the probability of ultimate ruin is given.
Furthermore, in the last of the paper a real data application will be shown.
Key Words: Matrix-exponential distribution, Phase-type distribution, Ruin probability, Markov chain.

Estimable Vectors and Models with Orthogonal Block Structure

Francisco Carvalho[1,2], João T. Mexia[1], Ricardo Covas[1,2]
[1]Centro de Matemática e Aplicações, Universidade Nova de Lisboa, [2]Unidade Departamental de Matemática e Física, Instituto Politécnico de Tomar, Portugal

In this paper we obtain a general expression for linear estimators of estimable vectors that are uniformly minimum variance unbiased estimators, when normality is assumed and, in general, best linear unbiased estimators, whatever the variance components, for models with orthogonal block structure. In general our estimators will be distinct from the least squares estimators. These models have as variance-covariance matrices all positive semi-definite matrices that are linear combinations of known orthogonal projection matrices, $\mathbf{Q}_1, \ldots, \mathbf{Q}_m$, that are pairwise orthogonal. When the orthogonal projection matrix $\mathbf{T}$ on the space spanned by the mean vector commutes with the matrices $\mathbf{Q}_1, \ldots, \mathbf{Q}_m$, it is shown that the new estimators are identical to the least squares estimators.
Keywords: Mixed linear models, OBS, Variance components, Commutative Jordan algebras.

References

1. Areia, A.; Carvalho, F. Perfect Families: an application to Orthogonal and Error Orthogonal Models. *AIP Conf. Proc. Numerical Analysis and Applied Mathematics: International Conference of Numerical Analysis and Applied Mathematics*, 1558, 841–846, (2013).
2. Bailey, R. A. Strata for randomized experiments. *Journal of the Royal Statistical Society, Series B*, 53, 27–78 (1991).
3. Brien, C. J. & Bailey, R. A. Multiple randomization. *Journal of the Royal Statistical Society, Series B*, 68, 571–599, (2006).
4. Caliński, T. & Kageyama, S. *Block Designs. A Randomization Approach. Volume I: Analysis*. Lecture Notes in Statistics. Springer, (2000).
5. Caliński, T. & Kageyama, S. *Block Designs. A Randomization Approach. Volume II: Design*. Lecture Notes in Statistics. Springer, (2003).
6. Carvalho, Francisco; Mexia, Jo˜ao T.; Santos, Carla. Commutative orthogonal block structure and error orthogonal models. *Electronic Journal of Linear Algebra* 25, 119–128, (2013).
7. Carvalho, F.; Mexia, J. T.; Oliveira, M. M. Canonic inference and commutative orthogonal block structure. *Discussiones Mathematicae* 28 (2), 171–181, (2008).
8. Ferreira, S.S.; Ferreira, D.; Fernandes, C.; Mexia, Jo˜ao T. Orthogonal models and perfect families of symmetric matrices. *Bulletin of the International Statistical Institute. Proc. ISI 2007, Lisboa 22–28 August*, 3252–3254, (2007).
9. Fonseca, M.; Mexia, J. T.; Zmy´slony, R. Inference in normal models with commutative orthogonal block structure. *Acta et Commentationes Universitatis Tartunesis de Mathematica*, 3–16, (2008).
10. Houtman, A. M.; Speed, T. P. Balance in designed experiments with orthogonal block structure. *The Annals of Statistics* 11 (4), 1069–1085, (1983).
11. Mejza, S. On some aspects of general balance in designed experiments. *Statistica* 52, 263–278, (1992).
12. Nelder, J.A. The Analysis of Randomized Experiments with Orthogonal Block Structure. I - Block Structure and the Null Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283 (1393), 147–162, (1965a).
13. Nelder, J.A. The Analysis of Randomized Experiments with Orthogonal Block Structure. II - Treatment, Structure and the General Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283 (1393), 163–178, (1965b).
14. Schott, James R. *Matrix Analysis for Statistics*, Wiley Series in Probability and Statistics, (1997).
15. Seely, J. Quadratic subspaces and completeness. *The Annals of Mathematical Statistics* 42, 710–721, (1971a).

16. VanLeeuwen, Dawn M.; Seely, Justus F.; Birks, David S. Sufficient conditions for orthogonal designs in mixed linear models. *Journal of Statistical Planning and Inference* 73, 373–389, (1998).
17. Zmyślony, R. A characterization of Best Linear Unbiased Estimators in the general linear model. *Lecture Notes Statistics* 2, 365–373, (1980).

Optimal Sample Size Allocation for Multi-Level Stress Testing with Exponential Regression Under Type-I Censoring

Chan P.s., Balakrishnan N., So S.Y., H.K. Ng
The Chinese University of Hong Kong

In this paper, we discuss the optimal allocation problem in a multi-level accelerated life testing experiment under Type-I censoring when an exponential regression model is used for the analysis. We derive the expected Fisher information matrix and use it to obtain the asymptotic variance-covariance matrix of the maximum likelihood estimators (MLEs). Then we consider the optimal allocation under the D-optimality criterion, and present an algorithm for determining the optimal allocation. A numerical example is presented for illustration. The optimal allocation depends on the model parameters and the sensitivity of the optimal allocation to misspecification of the model parameters is examined as well.
Key Words: Accelerated Life-testing; Fisher Information Matrix; Maximum likelihood estimators; Optimal design

Cohort effects and structural changes in mortality trend

Edviges Coelho[1], Luís Catela Nunes[2]
[1]Statistics Portugal, Lisbon, Portugal and Universidade Lusófona de Humanidades e Tecnologias, [2]Universidade Nova de Lisboa, Faculdade de Economia, Portugal

In Portugal, as in other developed countries, substantial mortality gains have occurred at all ages in recent decades. A exploratory study of the patterns of mortality decline in Portugal over the last sixty years shows that, although age-specific death rates have declined at all ages, the rhythm at which specific rates have been improving differs between ages and over periods of time. In particular, an apparently odd pattern of mortality improvement was detected in the male population, suggesting that over time a specific group of male individuals might have experienced a non-declining or even increasing mortality. Also, the presence of structural changes in overall mortality time trend in several

European countries (Coelho and Nunes, 2011), among them Portugal, has brought up questions about why such changes have occurred.
In this paper we focus on such questions. Specifically, we explore the presence of cohort effects as a possible reason behind the structural changes. We make use of visualizing techniques and several model specifications, namely the classical log-linear age-period-cohort (APC) model as well as more complex models with age-period and age-cohort interactions. A simulation study allows us to understand the impacts of ignoring important features in mortality modelling.
The paper is divided into two parts. In the first part, using post-1950 male mortality data for Portugal, we explore specific patterns in mortality that are consistent with the presence of cohort effects. We estimate several variants of the APC model. Data from England and Wales, where a cohort effect is well documented (Willets, 1999, 2004), is used as a benchmark. The impacts of using an incorrect specification of the model are also explored. In the second part, using the data and results from Coelho and Nunes (2011) for a set of 14 European countries, Canada, USA, and Japan, we analyse the possibility of an association between the presence of structural changes in the overall trend of mortality decline and the existence of cohort effects in the data.

References

1. Coelho, E. and Nunes, L. C. (2011). Forecasting mortality in the event of a structural change. Journal of the Royal Statistical Society A 174, Part 3, pp 713-736.
2. Willets, R.C. (1999). Mortality in the next millennium. Staple Inn Actuarial Society Meeting, 7 December 1999. Available at: http://www.sias.org.uk/.
3. Willets, R.C. (2004). The cohort effect: insights and explanations. British Actuarial Journal, 10, 833-877.

Adjustment Curves for Binary Responses Associated to Stochastic Processes

G.D. Costanzo, F. Dell'Accio, G. Trombetta
Università della Calabria, Italy

With regard to observed functional data and an associated binary response, we introduce a new method based on the definition of special random multiplicative cascades to simulate the underlying stochastic process. In particular, we consider a class S of stochastic processes whose realizations are real continuous piecewise linear functions satisfying a particular geometric condition. Let R be the family of all binary responses Y associated to a process X in S and consider data from a continuous phenomenon which can be simulated by a pair (X,Y) belonging to (S×R), we then introduce the adjustment curve for the

binary response Y of the process X that is a decreasing function which would make it possible to predict Y and then, if required, alter the process at each point in time before termination. For real industrial processes, which can be modeled by (X,Y) in (S×R) this curve can be a useful tool for monitoring and predicting the quality of the outcome before completion. Finally, the results of an application to data from an industrial kneading process are illustrated.

Key Words: Functional Data, Random Multiplicative Cascade, Adjustment Curve, Stochastic Process

Central limit theorems for an Indian buffet model with random weights

Irene Crimaldi
IMT Institute for Advanced Studies, Italy

The three-parameter Indian buffet process is generalized. The possibly different role played by customers is taken into account by suitable (random) weights. Various limit theorems are also proved for such generalized Indian buffet process. Let L_n be the number of dishes experimented by the first n customers, and let $\overline{K_n} = (1/n)\sum_{i=1}^{n} K_i$ where K_i is the number of dishes tried by customer i. The asymptotic distributions of L_n and $\overline{K_n}$, suitably centered and scaled, are obtained. The convergence turns out to be stable (and not only in distribution). As a particular case, the results apply to the standard (i.e., non generalized) Indian buffet process. The model investigated generally applies to evolutionary phenomena. In a biological framework, for instance, a new born exhibits some features in common with the existing units with a probability depending on the latter's weights (reproductive power, ability of adapting to new environmental conditions or to compete for finite resources, and so on). The new born also presents some new features that, in turn, will be transmitted to future generations with a probability depending on his/her weight. Similar examples arise in connection with the evolution of language. A neologism is often directly attributable to a specific people (or journal, period, event and so on) and its diffusion depends on the importance of such a people. Other applications of the weighted IBP could be found in Bayesian nonparametrics. Standard IBP is widely used as a prior on binary matrices with a fixed finite number of rows and infinitely many columns (rows correspond to objects and columns to features). The weighted IBP can be useful in all those settings where customers arrive sequentially. As an example, some dynamic networks present a competitive aspect, and not all nodes are equally successful in acquiring links. Suppose the network evolves in time, a node (customer) is added at every time step, and some links are created with some of the existing nodes. The

different ability of competing for links is modeled by a weight attached to each node. Each node could be described by a set of binary features (dishes) and the probability of a link is a function of the features of the involved nodes. A nonparametric latent feature model could be assessed at every time step, with the weighted IBP as a prior on the feature matrix. The main reference is

Berti P., Crimaldi I., Pratelli L., Rigo P. "Central limit theorems for an Indian buffet model with random weights". Available on ArXiv: 1304.3626v2 (2013). Submitted.

Keywords: Bayesian nonparametrics, Indian buffet process, Random reinforcement.

Modeling students' flow in higher education. The role of the membership function

Franca Crippa[1], Mariangela Zenga[2]
[1]Department of Psychology, University of Milano-Bicocca, [2]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy

Building the membership functions plays a key role in the applications of fuzzy sets theory, due to the fact that, depending on the applications, there are various ways to get reasonable membership functions. In educational studies, the issue has already been faced with respect grading systems, as Law (1996) introduced a method to build the membership functions for educational and other authors have proposed other advances (Weon & Kim, 2001).

With the introduction of Markov systems with fuzzy states (Symeonaki & Stamou) and their consequent application to Greek HE career paths (Symeonaki & Kalamatianou, 2011), the definition of the membership function acquires a crucial relevance in the analysis of students' retention, in the first place due to the existing differences in university systems. In our work, we discuss alternative membership functions for the Italian case.

References

Law, C. K., "Using Fuzzy Numbers in Educational Grading System", Fuzzy Sets and Systems,
Vol. 83, No. 3, 1996, pp. 311-324.

Learning achievement evaluation strategy using fuzzy membership function in: Frontiers in Education Conference, 2001. 31st Annual (Volume:1), pp.19-24, ISSN : 0190-5848 Print ISBN:0-7803-6669-7

Multistate model for evaluating conversion options in life insurance

Guglielmo D'Amico[1], Montserrat Guillen[2], Manca Raimondo[3]
[1]University G. d'Annunzio of Chieti-Pescara, Italy, [2]University of Barcelona, Spain, [3]Università di Roma "La Sapienza", Italy

In this paper we propose an age-time dependent semi-Markov model in discrete time to evaluate the conversion option in life insurance. This option is provided by life insurance companies extensively and allows the policyholder to convert the original policy into a different type of policy before the maturity of the former contract. The model focuses on the patient's age as a relevant factor to forecast the transitions among the different levels of seriousness of the disease and simultaneously on the chronological time.
Key Words: semi-Markov process, life insurance, conversion option

Is the Weibul distribution really suitable to forecast wind energy production?

G. D'Amico, F. Petroni, F. Prattico
Università degli Studi dell'Aquila, Italy

The importance of wind farms in the energy production is increasing year after year, but the energy produced by a wind farm has a stochastic nature mainly due to the variability of wind speed. Risk management techniques are to be used to take into account this variability when dealing with an investment in wind farm. In the present work we deal mainly with "operational risk" which is, in our opinion, of great importance. The sources of operational risk are linked to the distribution of wind speed that has to be well represented by a parametric distribution that takes into account rare events.
Here we show that the simple Weibull distribution underestimates rare events. We use than extreme value theory, through the Generalized Pareto distribution, in order to give a better estimation of the tail of the wind speed distribution.
Once the procedure is set up, we compare the results from the proposed procedure with those obtained by considering a simple Weibull distribution.
Key Words: Wind speed, Extreme value theory, Weibull distribution, Generalized Pareto distribution

Comparing Lee Carter extensions: efficiency vs accuracy

V. D'Amato[1], S. Haberman[2], G. Piscopo[3], M. Russolillo[1]
[1]Department of Statistics and Economics, Campus di Fisciano, University of Salerno, Italy, [2]Faculty of Actuarial Science and Insurance, Cass Business School, City University, UK, [3]Department of Economics, University of Genoa, Italy

The requirement of accurate mortality projections is becoming more pressing for both insurance companies and Authorities in order to capture, measure and manage longevity risk. For this aim, the most used mortality framework is the Lee Carter (LC) model (1992); recently, many extensions have been proposed to overcome the well know limits of the LC. In this work we propose a comparative assessment of two different approaches developed during the years we have dedicated to the topic. On the one hand, we have developed algorithms based on Variance Reduction Techniques to reduce the variability in the intervals of mortality projections; on the other hand we have introduced the dependence between different populations to improve the knowledge on the phenomenon. Of course, taking into account different sources of risks, like the dependence risk, the projection intervals becomes wider. We reflect upon these different points of view, in order to assess the consequences of the tradeoff between efficiency and accuracy in mortality projection on actuarial measures.
Keywords: Lee Carter model, mortality projections, population dependence, Variance Reduction Techniques.

References:

1. DAmato V., Haberman S., Piscopo G., Russolillo M.: Modelling dependent data for longevity projections. Insurance Mathematics and Economics 51, 694-701 (2012)
2. DAmato V., Haberman S., Piscopo G., Russolillo M: Computational Framework for Longevity, Risk Management, Computational Management Science (2013)
3. Lee, R.D., L. R., Carter: Modelling and Forecasting U.S.Mortality, Journal of the American Statistical Association, 87, 659-671 (1992)
4. Renshaw, A.E., Haberman, S.: A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factors, Insurance: Mathematics and Economics 38(3): 556-570 (2006)

Supercentenarians: the Exceptional Longevity by Extreme Value Theory

Valeria D'Amato[1], Tewfik Kernane[2]
[1]*Department of Economics and Statistics, University of Salerno, ITALY,*
[2]*Department of Probability and Statistics, Faculty of Mathematics, University of Sciences and Technology USTHB, ALGERIA*

The longevity improvements recorded in recent decades are pushing new limits as the proliferation of centenarians and supercentenarians, i.e. people 110 years old and older.
The number of supercentenarians in the International Database of Longevity (IDL) has been increasing exponentially since the mid-1970s, well documented particularly in Europe and Japan (Robine et al. 2002). Some authors, for instance Thatcher et al. (1998), have suggested that the longevity trend tends toward, never reaching a plateau or ceiling of mortality. The steady improvements of longevity phenomenon points out the challenge in mortality risk modeling, due to the importance in determining the survival probabilities at extreme ages. The longevity past experience could be misleading about the future evolution, since rare longevity events could become no longer few episodes of extreme mortality improvements. From risk management perspectives, the financial effect of underestimating the life table limiting age can be significant for insurance companies, pension providers, etc. The common risk measures implemented in the insurance industry to provide fundamental support for decision making such as VaR or Expected-Shortfall could lead to a significant bias, relying on the tail of the population survival distribution.
The Extreme value theory (EVT) can be adequately introduced as an accurate measure of the tail of the distribution, for evaluating extreme events and extrapolating the probability of even more extreme, out-of-sample events.
The recent literature on the problems related to mortality data at very old ages has shown an increasing interest towards the extreme value theory (EVT). The behaviour of distribution of human life spans has, among others, been discussed in Aarssen et al. (1994), Galambos et al. (2000) and Thatcher (1999) which determined the highest attainable age by using classical extreme value theory. For instance Beelders et al. (2004) valued the mortality bond using extreme value theory. They model the distribution of mortality by using a generalised Pareto distribution and price the Swiss Re mortality bond issued in 2003. In 2005 Han represent the trend of elderly people by employing EVT. Li et al. (2008) estimate survival distributions to extreme ages and the appropriate end point of a life table by EVT approach. In the longevity risk securitization context, Chen et al. (2010) represent the so-called rare longevity events exploiting the classical results of the EVT. In this paper we propose an

algorithm in the EM setting for studying the behaviour of mortality rates over a given high threshold age. A numerical application is provided.

References

1. Aarssen K., de Haan L. (1994). On the maximal lifespan of humans. Mathematical Population Studies 4, 259-281.
2. Baudisch, A. (2008). Inevitable Senescence? (Demographic Research Monographs; 4). Berlin[et al.]: Springer.
3. Beelders, O., Colarossi, D. (2004). Modelling mortality risk with extreme value theory: The case of Swiss Re's mortality-indexed bond. Global Association of Risk Professionals, 4:26-30.
4. Christensen, K., Johnson, T.E., and Vaupel, J.W. (2006). The quest for genetic determinants of human longevity: challenges and insights. Nature Reviews Genetics, 7:436- 448.
5. Chen, H., Cummins J.D. (2010). Longevity bond premiums: The extreme value approach and risk cubic pricing, Insurance: Mathematics and Economics, 46:150-161

NEW INTEGRAL RESULTS FOR RANDOM VARIABLES ESTIMATIONS

Zoubir DAHMANI
UMAB, University of Mostaganem, Algeria

In this talk, we present new results for random variables. We apply the fractional operators to establish new generalisations for expecations and variances for random variables defined on real intervalles. Some illustrative examples are also presented.

Some criteria to select a pricing measure for solving the valuation problem in incomplete markets

Silvia Dedu, Muhammad Sheraz, Vasile Preda
Department of Applied Mathematics, Bucharest University of Economic Studies, Faculty of Mathematics and Computer Science, Bucharest, Romania

In financial literature for incomplete markets there exist several equivalent pricing martingale measures. Therefore, the problem of selecting the most appropriate equivalent martingale measure is crucial and entropy measures play a very important role in this selection criteria. In this paper we present a comparative approach between Tsallis entropy, Kaniadakis entropy and entropies based on fractional calculus, relative to an incomplete market model. We consider the valuation problem in incomplete markets and analyze the empirical performance of

the methods proposed in case of some financial applications. Computational results are provided.
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.
Key Words: Tsallis entropy; Kaniadakis entropy; fractional calculus; incomplete markets; pricing measures.

Solvency capital within the Brownian framework partially modulated by a continuous-time Markov chain

Pierre Devolder, Adrien Lebègue
Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Belgium

A quite important issue in life insurance or pension benefits is the determination of the “right” or “fair” minimum level of capital needed by the insurer to be solvent, and especially under the future so called “Solvency II framework” in the European Union, which is designed to be more risk-sensitive. The particularity stands in the long-term characteristic of these products, and it may be important to take into account the time horizon effect of these products in the computation of the solvency capital, which is the purpose of this paper. We consider two market risks: the interest rate risk (with a Vasicek model) and the equity risk (with a Markov-modulated geometric Brownian motion, which incorporates a kind of economic risk with the continuous-time Markov chain which modulates the parameters of the geometric Brownian motion). Then, a portfolio is built on this market composed of three assets (a bank account, a zero-coupon bond and a stock) which satisfies the constant portfolio allocation hypothesis. The insurer offers a fixed guaranteed rate on a certain time horizon and invests the single premium in the portfolio. The probability of default is first considered, and then the issue of the determination of the solvency capital is addressed by the mean of the value at risk and the tail value at risk.
Keywords: solvency capital, Solvency II, risk measures, value at risk, tail value at risk, Vasicek, Markov-modulated geometric Brownian motion.

Fitting Binomial Distribution to Online Rating Data: TripAdvisor Ratings in Crete Island

Yiannis Dimotikalis
Department of Accounting & Finance, TEI of Crete, Greece

Last decade the growth of internet as a channel of promotion and trading of products and services, provide also the opportunity to buyers and users to review and rate their consumer experience. Online reviews and rates are very popular and companies like TripAdvisor and Yelp transform the user's product experience as the basis of their business model. Ratings usually collected and presented in scale of "five stars" rates and their statistical analysis is the subject of this paper. Because reviewers by their rate evaluate their likeness or dislikeness, we suggest and test simple Binomial distribution as a model of ratings and the Binomial's parameter p (probability of success) as the likeness parameter. To optimize the fitting of Binomial Distribution to real rating data, Maximum Likelihood Estimation (M.L.E.) of parameter p, is used as starting value of Least Square Error (L.S.E.) or Sum of Absolute Error (S.A.E.) minimization to find best p parameter value. For the fitting to rating data a suitable error measure, the false rate percentage, the percentage of reviewers who deviate from distribution estimations, is introduced and used as fitting criterion. Real datasets of TripAdvisor ratings of Crete island attractions and tourism sector services are used; to test Binomial distribution and proposed method of fitting performance. A similar size data set of Barcelona tourism sector ratings used as comparative benchmark. Also analysing hotel ratings data homogeneity of each rating dataset investigated. Finally we test for the effect and source of noise in data via simulation experiments. The main conclusion is that simple Binomial distribution with only one parameter, is capable to explain the actual distribution of online rating data with accepted error (noise or bias) in the majority of datasets.
KEY WORDS: Binomial distribution, Discrete Logistic distribution, online rating, TripAdvisor rating, simulation experiments, false rate criterion.

Approximating Inter-arrival and Service time Distributions by Phase-type Distributions in Single server Queues: A Strong Stability Approach

Yasmina Djabali, Boualem Rabta, Djamil Aïssani
LAMOS, Laboratory of Modelisation and Optimization of Systems, University of Béjaia, Algeria

In our work, we provide by means of the strong stability method, the mathematical justification of the approximation method by phase type

distributions that is already used in several works. We consider the approximation of GI/M/1(resp.M/G/1) queueing system by a PH/M/1(resp.M/PH/1) system, where PH refers to a hyperexponential H2 or a hypoexponential HOE2 distribution depending on the value of the coefficient of variation of the original distribution. Approximations are compared with the simulations.

Key Words: Queueing systems, Phase-type distributions, Perturbation, Strong stability, Markov chain.

A Novel Approach to the study of the EEG Signals Scaling Properties in Various States of Consciousness

Ludmila A. Dmitrieva[1,2], Darya A. Zorina[2], Igor E. Kanunikov[2,3], Maria N. Krivoshapova[4], Yuri A. Kuperin[1,2], Maria A. Shaptiley[4]

[1]Department of Physics, Saint-Petersburg State University, [2]Department of Liberal Arts and Sciences, Saint-Petersburg State University, [3]Department of Biology, Saint-Petersburg State University, [4]Scientific Research Center of Health "Exiclub", Russia

In this work to find the difference between the states of rest and meditation we study the scaling properties of the corresponding multichannel EEG for a group of subjects. Namely, we calculate the so-called self affine index. This index is defined as the exponent in power law dependence of the sum of the EEG samples on the amount of items. The self affine index has been calculated for 20 fragments consisting of 1000 samples of each EEG channel for the group of subjects. This group consisted of five experienced in the practice of meditation subjects and ten inexperienced subjects. EEG have been recorded in the rest state and during meditation. The latter was carried out by the method of «Open Monitoring» classification by Lutz and Davidson. This technique is aimed to minimize the spontaneous cascade of semantic associations. We then statistically processed the obtained results by means of repeated ANOVA. The main results are the following. For experienced in the practice of meditation subjects the self affine index in a state of meditation was significantly higher as compared to the same index at rest state. For inexperienced subjects difference in self-affinity index at rest and a state of meditation was not statistically significant. For all subjects in meditation state the self affine index was significantly higher in the left anterior frontal region. Thus it can be assumed that the higher self-affine index in the left frontal region reflects higher compared to other channels synchronization of sources involved in the generation of the EEG. More detailed neurophysiologic explanations of the obtained results will be presented in the talk.

Keywords: self affine index, multichannel EEG, scaling, meditation, method of «Open Monitoring», statistical tests.

*The work is supported by the SPbU grant #37.23.1496.2013

The Study of Correlation Dimension of the EEG Signals in a State of Meditation by means of Empirical Mode Decomposition

Ludmila A. Dmitrieva[1,2], Igor E. Kanunikov[2,3], Maria N. Krivoshapova[4], Yuri A. Kuperin[1,2], Nikolai M. Smetanin[1], Maria A. Shaptiley[4]

[1]Department of Physics, Saint-Petersburg State University, [2]Department of Liberal Arts and Sciences, Saint-Petersburg State University, [3]Department of Biology, Saint-Petersburg State University, University Embankment, [4]Scientific Research Center of Health "Exiclub", Russia

The aim of this work was to develop an indicator for finding differences in the multichannel EEG of experienced and inexperienced subjects in the states of meditation and rest. In our study, as such an indicator, we have chosen the difference between the correlation dimensions of the EEG channels reconstructed attractors and the corresponding EMD-filtered correlation dimensions (we denote this difference as DifD2). Under the EMD-filtered EEG correlation dimension here we understand the correlation dimension calculated from EEG signal, in which the first two modes of empirical decomposition (EMD) are dropped. By means of the method of local divergence rates of the reconstructed attractors for EEG signals, the authors previously showed that the sum of the first two EMD-modes of EEG are stochastic components (both physical and physiological noise). Thus, the smaller the difference DifD2, the less noise in the EEG signals. Processing algorithm of EEG database for both experienced and inexperienced subjects was the following. Calculations were performed for 5 experienced and 5 inexperienced subjects in the states of meditation and rest. For each subject we calculated the difference DifD2 over 10 fragments of each channel of multichannel EEG (19 channels). Each fragment consisted of 10,000 samples of EEG signal. Unfiltered correlation dimensions were calculated in the embedding dimension equal to 5. After dropping the first two modes, EMD-filtered correlation dimensions are calculated in the embedding dimension equal to 4. Then the obtained results of calculations were processed by standard statistical method "Repeated ANOVA". It turned out that the DifD2 in a state of meditation for experienced subjects was significantly lower than that of the inexperienced subjects. It should be noted that this distinction is achieved mainly due to the frontal and parietal channels. For experienced subjects the difference DifD2 in the meditation state turned out to be significantly lower than that in the rest state. It should be noted that this difference is achieved by the parietal and occipital channels. At the same time, the difference DifD2 in meditation and in the rest states disappears for inexperienced subjects. Neurophysiologic explanations of the obtained results will be presented in the talk.

Keywords: correlation dimension, empirical mode decomposition (EMD), EMD-filtered correlation dimensions, meditation, indicator DifD2.
The work is supported by the SPbU grant #37.23.1496.2013

The multi-dimensional development of developing countries: trends and country profiles for 2000-2010 period

Céline Dondeynaz[1,2], César Carmona Moreno[1]
[1]*Insitute for Sustainability and Environment, Joint Research Centre, Italy,*
[2]*School of engineering, Liverpool University, UK*

Water Supply and Sanitation (WSS) services assume a similar key role if considering human/society development: water supply heavy burden and lack of sanitation slows down the economy, social improvement and population well-being. They also interact with the environment, the water resources and the way they are managed. As focusing on the developing world, the Official Development Assistance (ODA) in recipient countries is included in the analysis and modelling (Dondeynaz et al, 2013). The approach proposed in this paper is global targeting cross-countries analysis and comparison to capture trends of water supply and sanitation coverage and its benefits for country development. These research outputs could support national decision making and/or donors 'strategies
The Joint Research Centre – European Commission has developed a coherent database (WatSan4Dev) that includes the most influential variables on WSS services in developing countries (official indicators provided by international agencies). The 25 variables for 101 developing countries selected are organized into five thematic areas: Water Resources-WR, human Activity Pressure on water resources-AP, Country Environmental Concern-CEC, Human Development and Poverty –HDP and Official Development Aid-ODA (Dondeynaz et al, 2012). This research has built on these 5 thematic axes, five country profiles in 2004 (one-year first step). The k-mean clustering method was used to identify groups of countries with similar profiles among the 101 country observations.
Country profiles were ordered from most favourable (profile 1) to the less favourable country situation (profile 5). Profile 1 presents relatively equilibrated and high values on WR, HDP and CEC implying little need for external support (ODA). Profile 2 shows weaknesses in terms of accountability and population's freedom, leading to a low commitment towards environment (CEC). Profile 3 indicates that the economy is mainly driven by agricultural activities facilitated by a context of abundant natural resources (AP). Profile 4 and 5 are the less favourable profiles when considering human development (HDP). This classification is a new way to identify countries in difficulty but also indicates weaknesses restraining countries' development and well-being.

This paper will present the extension and the methods used to analyze trends over the last decade (2000-2010). The analytical framework allows observing the evolution in ranking of the developing countries selected and map the changes of clusters over 2000, 2005, 2010. In particular it will be observed the evolution of countries benefiting from high ODA, if there is a progress towards best and equilibrated profile (Profile1). The methodological process to assess trends and profile changes during the last decade will be detailed and presented.

Keywords: Multivariate analyses, clustering methods, composite indicator, longitudinal data analysis, water supply and sanitation, developing countries

The Determinants of Demand for Health Insurance in Ghana: Does Perception of Quality of Healthcare Services Matter?

Stephen K. O. Duku[1, 2, 3], Edward Nketiah-Amponsah[4], Wendy Janssen[2], Menno Pradhan[2]

[1]Noguchi Memorial Institute for Medical Research, University of Ghana, Ghana, [2]Faculty of Economics and Business Administration, Development Economic, Free University of Amsterdam, The Netherlands, [3]Amsterdam Institute for Global health and Development, The Netherlands, [4]Department of Economic, University of Ghana, Ghana

Background

Some sub-Saharan African counties such as Gabon, Mali, Senegal, Rwanda and Ghana have taken giant steps towards universal health coverage by adopting prepayment and risk pooling systems to provide financial protection particularly to the poor and vulnerable in society. Ghana implemented the National Health Insurance Scheme (NHIS) in 2004. As at December 2011, the NHIS has over 8 million active membership representing 33% of the Ghanaian population (NHIA, 2012). Despite this achievements, recent research has identified some implementation challenges such as delays in the production and distribution of NHIS ID cards, long waiting times at health facilities, poor health staff attitude, inadequate information on NHIS benefit package, delays in reimbursing health providers, perceived poor quality of NHIS and health provider services and under-the-table payments to care providers (Jehu-Appiah et al. 2011; Aikins & Dzikunu, 2006; Sulzbach et al. 2005). This study aims at assessing the determinants of demand for health insurance, with specific focus on whether perceptions on quality of health provider services matter. The findings from this study are expected to inform policy makers on areas to focus interventions to increase enrolment and renewal in the NHIS. It will also be of immense value to other Sub-Saharan African countries implementing social health insurance.

Methods

We used primary data from a sample 4,214 individuals 18 years and above living within 5 – 10 km radius of the catchment area of 64 selected primary health care facilities in 16 districts in the Greater Accra and Western regions of Ghana. Bivariate analysis and probit regression estimation reporting marginal effect was employed to examine the relationship between the perception of quality of health provider's services and health insurance enrolment. The regression estimations were done separately for each of the three quality perception variables. Other theoretically relevant socio-economic variables were also included in the regression estimation.

Results

We found no significant difference between the perceptions of the insured and uninsured on non-technical quality of health provider services. However, the marginal effects indicates that generally, individuals with positive perceptions of the non-technical quality of care have a higher percentage points probability of enrolling in health insurance than those with negative perceptions for all three non-technical quality variables. Our study found socio-economic variables such as age, marital status, educational level, sector of employment, household size, locality of residence, health status and wealth status to be stronger determinants of demand for health insurance in Ghana.

Conclusion

From our findings, we argue that clients perception of the quality of health services matter in the health insurance enrolment decision in Ghana. We further argue that clients attach great importance to the non-technical quality and interpersonal relation with services providers such as staff attitude, waiting time and fair queuing system. We suggests that any attempt at address non-technical quality attributes should take into consideration measures aimed first at identifying these context specific important quality attributes before designing interventions that are culturally and socially acceptable in the local or regional context. Whiles policy makers strive to improve the non-technical quality of services at accredited health facilities, they should not loose site of the socio-economic factors that have stronger influence on the enrolment decision and the need to continuously support poor and disadvantaged families to enroll in the NHIS as well as encourage the rich and healthy in society to enroll through information provision and mass registration campaigns.

Key words: Determinants, Health Insurance, Quality of healthcare, National Health Insurance Scheme, Ghana

Bootstrap and Gibbs Sampler Of A Parametric Markov Random Field On Image Processing

Abdeslam El Moudden
Faculté Polydisciplainre, Université Cadi Ayyad Marrakech, Morocco

In this paper, we treat the problem of parameter estimation of the energy function associated to a Markov random field. It is known after the Kolmogorov theorem (Billingsley, L. 1979) that the existence and uniqueness of a probability measure of a Markov random field depends on the existence of a family of projective measures. Thus, such a family of projective measures can be represented by a parametric family function potential which summarizes the nature of interactions between pixels in a network. Therefore, solving the problem related to the estimation of such a parametric model requires a more accuracy. We introduce the bootstrap (Efron, B. and Tibshirani, R. 1993) and the Gibbs sampler (Besag, J. 1995) elementary or per block to retrieve the probability measure. The consistency of the bootstrap estimate is indeed proven in this context.
Key Words: Markov Random Fields, Gibbs Sampler, Bootstrap, Parametric Estimation, Image processing.

BINOMIAL TREE MALLIAVIN CALCULUS AND RISK MEASURES

Robert Elliott[1], Tak Kuen Siu[2], Sam Cohen[3]
[1]University of Adelaide and UNISA, Canada, [2]Cass Business School, City University London, UK, [3]Mathematical Institute, Oxford, UK

The classical familiar framework used to introduce financial pricing is the binomial model. The talk will discuss several more advanced concepts in this simple framework. These will include martingale representation, Malliavin derivatives, backward stochastic difference equations and dynamic risk measures. The latter are introduced using non linear expectations which are the solutions of backward stochastic difference equations.
Key Words: backward stochastic equations; Malliavin calculus, binomial models, risk measures

Non-normalized PageRank and Random walks on N-partite graphs

Christopher Engström, Sergei Silvestrov
Mälardalen University, Sweden

In this article we will look at a variation of the PageRank algorithm originally used by L. Page and S. Brinn to rank home pages on the Web. We will look at non-normalized variation of PageRank and show how this version of PageRank relates to a random walk on a graph. The article has its main focus in understanding the behavior of the ranking as the system is expanding or contracting by adding or removing vertices or edges in the graph. More specific we will look at N-partite graphs and see that by considering the related random walk on the graph we can find explicit formulas for PageRank of the vertices in the graph. Last we will take a look at some small local changes in the graph as well as changes in personalization vector V and draw some conclusions from our results.
Key Words: PageRank, Markov chains, Random walk on a graph, N-partite graph

Genetic algorithm-based tuning of the *C-Value* for term ranking

Christopher Engström[1], Thierry Hamon[2;3], Sergei Silvestrov[1]
[1]Division of Applied mathematics, School of Education, culture and communication, Mälardalen University, Sweden, [2]LIMSI-CNRS, France, [3]University Paris 13, France

Text mining in scientific and technical fields requires terminological resources to access the knowledge of the domain, but such resources suffer of low coverage. Approaches based on linguistic rules have been proposed to automatically extract terms to help the terminology building from corpora. However, the quality of the term extraction results is not sufficient and existing term ranking metrics fail to offer convicting results. We propose to improve the *C-Value* ranking metrics by considering the syntactic role of the nested terms and by optimising the parameters with a genetic algorithm. Evaluation performs on a biomedical text collection demonstrates the *C-Value* parametrisation better rank the extracted terms: the average precision increases by 9% when compared to the frequency based ranking and 12% when compared to the *C-Value* based ranking.
Keywords: *C-Value*, Genetic algorithm, Terminology, Text Mining, Natural language processing.

Enumeration and classification of small orthogonal latin hypercubes for computer experiments

Haralambos Evangelaras
Department of Statistics and Insurance Science, University of Piraeus, Greece

In this paper we construct and enumerate all inequivalent orthogonal latin hypercubes with $n < 16$ runs and explore their space filling properties. Best designs are identified with respect to several selection criteria.
Keywords: Latin hypercube, orthogonality, uniformity, computer experiments

A model based criterion for the selection of a proper design for factorial experiments

Haralambos Evangelaras, Christos Peveretos
Department of Statistics and Insurance Science, University of Piraeus, Greece

In this paper we propose an effective procedure for the selection of a proper factorial design. This procedure is based on a simulation study, where competitive designs are tested for their capability in correctly identifying a given true model. Detailed examples are given for the classification of the non isomorphic two level orthogonal arrays OA(n,k,2,t), which can be used to study $3 \leq k \leq 6$ factors with $12 \leq n \leq 28$ experimental runs.
Keywords: Orthogonal arrays, Design selection, Estimation capacity, Simulation, False Discovery Rate

Fitting mixtures of linear mixed models: a simulation study

Susana Faria
University of Minho, Department of Mathematics and Applications, Portugal

Finite mixture models are a well-known method for modelling data that arise from a heterogeneous population. Within the family of mixture models, mixtures of mixed effects models have also been studied extensively.
Finite mixtures of mixed effects models conveniently allow to account for correlations between observations from the same individual and to model unobserved heterogeneity between individuals at the same time.
In this paper we propose to compare two algorithms to compute maximum likelihood estimates of the parameters of a mixture of linear mixed models. The comparison of the two procedures was done through

a simulation study of the performance of these approaches on simulated data sets.
Key Words: Maximum likelihood estimation; mixtures of linear mixed models; Simulation study

A Taylor Series Functional approximation to the Numerical Analysis of a non Periodic Review (S,s) Inventory Model

Rahmoune Fazia
Lamos Laboratory- University of Bejaia, Algeria

Real life inventory problems are often very complicated and they are resolved only through approximations. Therefore, it is very important to justify this approximation and to estimate the resultant error. This paper presents a functional approximation of the non periodic review inventory model with an (S,s) policy, built on a Taylor series approximation. Using the underlying Markov chain with respect to the perturbation of the demand distribution, we obtain quantitative estimates with an exact computation of constants. Numerical examples are carried out to illustrate the performance of our approach.
Key Words: (S,s) Inventory Model, Taylor Series, Markov Chains, Finctional Approximation, Numerical Analysis

Multivariate Markov chain predictions adjusted with copula models

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

In this paper we propose an estimator for the joint probability of Markov chains of finite order, with finite alphabet. We base the estimator's construction on the Partition Markov models (see [1] and [2]) which are generalizations of Variable Length Markov Chains. The technique can be summarized in two steps. First, the joint probabilities are found through a Partition Markov Model, using the Bayesian Information Criterion. Let us denote this estimator by P1. Second, the discrete copula related to P1 is computed (for discrete copula models [3] can be consulted) and the Partition Markov estimation is applied to each marginal process. Then, the estimated copula is rearranged in the space defined by the marginal estimations and a new estimator, P2, is defined. The efficiency of this new estimator is tested by simulations studies obtaining that P2 generates more accurate estimates than P1 even when the sample size is moderate. In addition, P2 offers a more robust estimation than P1 even when the order of the underlying Markov process grows. Finally, the use of the methodology is exemplified by jointly modeling two

financial discrete time series representing daily high, medium and low returns of two different assets.

Key Words: Multivariate Markov chains, Copula theory, Model selection, Correction through the marginal, Financial market dynamics.

References

[1]García, J. E. and Fernández, M.. Copula based model correction for bivariate Bernoulli financial series, AIP Conference Proceedings, 1558, 1487–1490 (2013).

[2]García, J. E. and González-López, V. A. Minimal Markov models. arXiv:1002.0729v1 (2010).

[3]Joe H. Multivariate models and dependence concepts, Monographs on Statistics and Applied Probability, vol. 73. Chapman & Hall/CRC: New York, (2001).

A heuristic procedure to estimate the tail index

Marta Ferreira

Center of Mathematics of Minho University and Department of Mathematics and Applications of Minho University, Portugal

The estimation of the tail index is a very important issue within extreme value theory. Semi-parametric estimators usually require the choice of the number k of upper order statistics to use in the estimation, which is a problem difficult to handle. Here we apply a heuristic graphical method to well-known semi-parametric tail index estimators and analyze it through simulation. We shall see that some criteria lead to good results.

Key Words: extreme value theory, tail index

Estimation in models with Orthogonal Block Structure

Sandra S. Ferreira[1], Célia Nunes[1], Dário Ferreira[1], João Tiago Mexia[2]

[1]Department of Mathematics and Center of Mathematics, University of Beira Interior,[2]Department of Mathematics and Center of Mathematics and Applications FCT/UNL, Portugal

We intend to present estimators with good behaviors for estimable vectors and variance components for a class of models that contains the well known models with orthogonal block structure, OBS. These estimators uses commutative Jordan Algebras and extends the one given for a more restricted class of models, the models with commutative orthogonal block structure, COBS, in which the orthogonal projection matrix on the space spanned by the means vector commute with all variance-covariance matrices.

Keywords: COBS, Commutative Jordan algebras, LSE, Orthogonal block structure, UMVUE.
This work was partially supported by the Center of Mathematics, University of Beira Interior through the project PEst-OE/MAT/UI0212/2014 and by CMA, Faculty of Science and Technology, New University of Lisbon, through the project PEst-OE/MAT/UI0297/2011.

References

1. Caliński, T., Kageyama, S., Block Designs: A Randomization Approach, Volume I: Analysis, Lecture Notes in Statistics, 150 Springer, New York (2000)
2. Caliński, T., Kageyama, S. Block Designs: A Randomization Approach, Volume II: Design, Lecture Notes in Statistics, 170, Springer, New York (2003)
3. Carvalho, F., Mexia, J.T., Oliveira, M., "Canonic inference and commutative orthogonal block structure". Discussiones Mathematicae. Probability and Statistics, 28, 171–181 (2008)
4. Ferreira,S.S., Ferreira, D., Nunes, C. and Mexia, J.T., "Estimation of Variance Components in Linear Mixed Models with Commutative Orthogonal Block Structure". Revista Colombiana de Estadstica 36, (2) 261–271 (2013)
5. Fonseca, M., Mexia, J.T.,Zmy´slony, R., "Inference in normal models with commutative orthogonal block structure". Acta et Commentationes Universitatis Tartuensis de Mathematica, 12, 3–16 (2008)
6. Mexia, J.T., "Best linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in the presence of controlled heteroscedasticity". Compututational Statistics & Data Analysis, 10 (3) 271–281 (1990)

Monitoring the variability of a multivariate normal process using STATIS

Adelaide Maria Figueiredo[1], Fernanda Otília Figueiredo[2]
[1]Faculty of Economics of University of Porto and LIAAD-INESC Porto, [2]Faculty of Economics of University of Porto and CEAUL, Portugal

In real situations, the evaluation of the global quality of either a product or a service depends on more than one quality characteristic. Often the quality characteristics of interest are correlated and so multivariate techniques of process control are more appropriate than univariate methods for monitoring the individual characteristics. In a multivariate quality control procedure is essential to identify the out-of-control variables when the statistical process control gives an out-of-control signal. In order to monitor multivariate processes and identify the variables responsible for changes in the process, we shall use the

STATIS (Structuration des Tableaux a Trois Indices de la Statistique) methodology – a three-way data analysis method.
Keywords: Process monitoring, *RV* coefficient, STATIS, Statistical Quality Control.

Resampling methodologies in Phase I control charts

Fernanda O. Figueiredo[1], M. Ivette Gomes[2]
[1]FEP, Porto University, and CEAUL, [2]FCUL, DEIO and CEAUL, Lisbon University, Portugal

The control charts, introduced by Shewhart in 1924, are one of the main tools in SPC, but their domain has been successively enlarged, with applications to areas as diverse as Health, Medicine, Genetics, Biology, Environmental Sciences, Finance, Sports and Justice, among others. To developany control chart the nominal process parameters must be either assumed known or estimated. In practice the distribution of the process data as well as the process parameters are usually unknown, being the process parameters usually estimated from an in-control historical data set (called a Phase I reference sample, being made up of m subgroups of size n) before proceed to the building of a (non-)parametric control chart.
A strong emphasis has been given to the analysis of the real performance of control charts implemented on the basis of estimated parameters. If we want to obtain control charts with estimated control limits with the same run-length properties as the corresponding charts with true limits, the choice of the number of subgroups m, and the sample size n, cannot be heuristic. It is commonly referred in the literature that we need a very large number m of subgroups (which is a limitation from a practical point of view, and in some cases, even impossible) and that we must determine the control limits in a robust way.
Our aim in this paper is to investigate the benefits of using resampling methodologies (in particular, the bootstrap) in Phase I of control charting to estimate the nominal process parameters, together with the use of robust estimates. More precisely, from a reference sample of m subgroups (20 or 30) of size n (4 or 5), we set out to obtain a larger reference sample obtained via bootstrapping to estimate the process parameters, and then, analyze the performance of the obtained control charts implemented on the basis of these estimates.

A quantile regression approach to male and female wage function in Portugal. Does the database matter?

Maria da Conceição Figueiredo[1], Elsa Fontainha[2]
[1]BRU-ISCTE-IUL, ISCTE – IUL Lisbon University Institute, [2]Department of Economics, ISEG – University of Lisbon, Portugal

The research aims to study the distribution of hourly wages for men and women in Portugal, adopting a quantile regression (QR) approach. Two databases are used for the estimation of the wage functions: the Quadros de Pessoal, Linked Employer-Employee Data (QP-LEED, administrative microdata) and the Inquérito ao Emprego, Portuguese Labour Force Survey (IE-LFS, Eurostat microdata).
Three basic models are considered to explain the hourly wages for men and women: the first model, using each database separately, is estimated adopting education, tenure, potential experience, activity sector, and job as independent variables; the second, using data from QP-LEED, includes additional determinants related to firm (firm size and foreign social capital); and the third, using data from the IE-LFS, includes additional independent variables related to the worker's family (marital status and children).
The results indicate that: (i) Regardless of the database used, the quantile regression (QR) shows superiority over OLS approach; (ii) In general, the same model specification estimated using each database - one administrative (QP-LEED), and the other based on a survey (IE-LFS) - present convergent results; (iii) Independently of the database used, the equations for men and for women reveal that the levels of education have a higher impact on wage determination; (iv) In general, the variables related to the firm contribute to the explanation of wages of men and women while those related to family only contribute to the explanation of men's wages; and (v) the clear different returns for the same characteristics found between men and women, and the pattern of differences which increase across quantiles strongly indicates that the present study should continue in the future, with the analysis of the explanation of the gender wage gap.
Keywords: wage functions; quantile regression; Linked Employer-Employee Data; Labour Force Survey; male-female wage differences

Parameters Estimation Methods for the "Enforced Regression" Problem

Jerzy K. Filus[1], Lidia Z. Filus[2]
[1]Department of Mathematics and Computer Science, Oakton Community College, [2]Department of Mathematics, Northeastern Illinois University, USA

Using the method of parameter dependence we describe stochastic dependence of a random variable, say Y, on a set of explanatory variables $X_1, \ldots, X_k$ as the conditional survival function $P(Y \geq y \mid x_1, \ldots, x_k)$. The relative ease of construction of these conditional distributions when using our "method of parameter dependence", prompted us to apply these distributions as an alternative to existing regression theories. This alternative relies on replacing the, typical, conditional expectation $E[Y \mid x_1, \ldots, x_k]$ 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 the accuracy of the existing dependence models. It also essentially differs from the "Regression Quantiles" concept by Koenker and Basset (1978). The problem that arises with our models is finding proper methods for estimating the numerical values of the underlying "parameter functions'" unknown coefficients. Two methods for solving this problem will be discussed.

Annuities calculation in Algeria: Continuous time approach

Farid FLICI
Research Centre in Applied Economics for Development, Statistics and Applied Economics, Higher School of Statistics and Applied Economics, Algeria

The aim of the present paper is to show and quantify the impact of calculation in continuous time in matter of annuity pricing and provisioning compared to discrete time calculation. The discrete approach allows insurance companies to facilitate calculation process but with less accuracy. This approach implies constancy of death quotients during a year. The people that were born in the same year are considered exposed to the same death risk. If the hypothesis of interest rate stability cannot be taken given the relative stagnation of the Algerian financial market, the demographic hypothesis has an important effect on annuities calculation. Moreover, insurance pricing is based on the equivalence between premium and annuity considering financial and demographic aspects. For more accuracy, this equivalency needs to take into account the exact dates of different monetary flows: premium and annuity. This is only possible by the continuous approach.
The use of this method allows more accuracy then more equity in annuities pricing in the way that for each risk it allows to give more

adequate premium and more flexible calculation. However, the duration and the date of accounting, perfection is the same with less hypothesis.
Annuities calculation is based on two essential parameters: interest rate and life table. With continuous approach, these parameters should be reformulated in continuous form. If i indicates the annual interest rate used, the instantaneous interest rate is given by r=ln (1+i). Similarly, life table is replaced by mortality law. To do this, several mortality models can be used: Makeham, Gompertz, Exponential…etc. the choice of the adequate model is obtained by Least Square Errors between table values and model values. For mortality in old ages, specific models are used. The mortality changes trend beyond specific age [Caole & Gou 1990], usually 80 or 85. Numerous models can be implemented for [Quashie & Denuit 2005]. The Data used in this paper is the Algerian life tables (Male and Female) published by the National Statistics Office for the years 20010, 2011and 2012. All these tables are extended to the age of 85. In first, average table is constructed. The second step is to moving from the five-year table to the single year table by using interpolating techniques [Lagrange, Karup-King]. Then, we use fitting mortality models to estimation the parameters of the mortality function. Beyond the age of 80, we have to calculate the adequate function.
Finally, we calculate the mathematical provision by using the same interest rate and the same life table but with different approaches: discrete and continuous. For this, we use the insured portfolio of the National Insurance Company (SAA – Algeria). This portfolio is composed of 1700 pensioners (supplementary pension). The premium and the provision given by the continuous approach are reduced comparatively with which was given by the discrete approach with about 7%. Tariffs discount is able to bring more attractiveness for the product without affecting the financial equilibrium of the insurance company.

Estimation of the missing life tables in the Algerian mortality surface (1977 – 2012) by using Lee Carter Formula

Farid FLICI
Research Centre in Applied Economics for Development, Statistics and Applied Economics, Higher School of Statistics and Applied Economics, Algeria

The construction of prospective life tables, the population projections and the analysis of mortality evolution need to make available continuous mortality surface. In Algeria, the first life table based on the civil registration data was been constructed in 1977. At that time, because of the shortage of data and experience, the life tables could not be constructed with annual frequency. This became possible from 1998. However, no one has tried to estimate the missing life tables for the previous period. Today, it is very necessary to make available this data,

especially for the governmental previsions and the life insurance calculations. This is the principal objective of the present paper. Several methods can be used to estimate the missing life tables. Because of peculiarity of the Algerian data, the set of the methods that can be used is restricted. Lee Carter Formula [1992] is initially developed to project the mortality age structure in the future. But before, the mortality surface should be fitted. The idea is to estimate the parameters of Lee Carter Formula by using available data, and to use the obtained formula to estimate age mortality structure for the missing years, which comprise life table.

Keys-words: Life table, Lee- Carter, Fitting, Missing Data, Algeria.

Number of patients at Kavala's hospital from 2005 until 2011

FLOROU G., AGGELIDIS V., BATZIOS C., KARASAVVOGLOU A., PETASAKIS I., POLYCHRONIDOU P.
Kavala Institute of Technology, Greece

In this paper we analyze the number of patients at each department at Kavala's hospital last years. Kavala is a city in the North Greece. We study for its department the patients' number per year, gender, days of stay at hospital, the insurance organism which pays that cost and the hospital ratios as average day of hospitalization, percentage of occupation, etc. We use real data from Kavala's hospital for years 2005 until 2011. We present the patient number of each department for each year and we analyze the decrease or increase of this number in the years of economic crisis.

We present the differentiations of hospital ratios in the years 2005 until 2011 also. We analyze the insurance organism of patients, as well as the number of patient without insurance organism, or patients without incomes.

We use descriptive statistic methods and parametric methods to test our hypotheses.

Key Words: Hospital, Greece, hospital ratio, patient

BIBD, Hadamard Matrices and Combinatorial Analysis

Carla Francisco[1], Teresa A.Oliveira[1,2], Donald Giles[3,1]
[1]Departamento de Ciências e Tecnologia, Universidade Aberta, Portugal
[2]Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal
[3]Department Head Norfolk Board of Education, USA

The foundations of block designs were laid down by R.A. Fisher in the beginning of 20th century with applications in agriculture. It became one

of the areas of mass development for research in various fields, from agriculture to medicine, playing a central role not only on the research process but also on the new technological advances, as we will see.
As a branch of block designs, research on Balanced Incomplete Block Designs (BIBD) arose several interesting and defying problems within combinatory mathematics. We will review some of the main concepts in BIBD and we will use skew designs to define skew balanced incomplete block designs based on their incidence matrix and their relation to Hadamard skew matrices. We will present how to construct some skew Hadamard matrices, and we will illustrate the algorithm with an application.
Hadamard matrices are present in our daily life, it is easy and current to find their different applications based on new technologies and codes of figures such as Quick Response (QR) codes. QR codes are bi-dimensional bar codes that can be easily read by common devices which have image capitation function, such as mobile phones. The use of such codes is very popular nowadays in simple things like to send a simple text message, a picture or for a batch of information regarding a product or whatever one desires. Connections between the presented methodologies and Quick Response (QR) codes will be discussed.
Keywords: BIBD, Experimental Design, Skew design, Hadamard Matrix, QR Codes Ackowledgments: This research was partially sponsored by national founds through the Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT under the project PEst-OE/MAT/UI0006/2014.

References

Baumert, L. D., S. W. Golomb and M. Hall, Jr. (1962). Discovery of an Hadamard matrix of order 92. Bull. Amer. Math. Soc. 68, 237-238
Dragomir Z. Djokovic, Oleg Golubitsky, Ilias S. Kotsireas (2013). Cornell University Library 2 Oct 2013.
Georgiou, S., Koukouvinos, C. Seberry, J. (2002). Hadamard matrices, orthogonal designs and construction algorithms, pp. 133-205 in DESIGNS: Further computational and constructive design theory, Kluwer 2003.
Hall, M., Jr.(1986). Combinatorial Theory, 2nd ed. New York: Wiley.
Hedayat, A. and Wallis, W. D. (1978). Hadamard matrices and their applications. The Annals of Statistics, Vol. 6, 1184-1238
Koukouvinos, C., Stylianou, S.(2008). Combinatorial Designs: A tribute to Jennifer Seberry. Volume 308, Issue 13, 6 July 2008, Pages 2723–2731
Mathon, R., Rosa, A.(2010) – CRC Handbook of Combinatorial Designs.
Oliveira, T. A. (2010). BIB designs with repeated blocks: review and perspectives, The Islamic Countries Society of Statistical Sciences, Vol. I, 82-96.
Plackett, R. L. and J. P. Burman (1946). The design of optimum multifactorial experiments. Biometrika 33, 305-325.

Raghvarao, D. (1971). Constructions and Combinatorial Problems in Design of Experiments. New York: Wiley.
Seberry. J. (2012). "Hadamard matrices" University of Wollongong. Available at http://works.bepress.com/jseberry/370.
Stanton, R.G., Kalbfleisch J. G. and Mullin, R. C. (1970). Covering and packing designs.In:Proceedings of the Second Chapel Hill Conference on Combinatorial Mathematics and its Applications, pp. 428-450. University of North Carolina, Chapel Hill, NC, 1970.

Population Health Index in Portugal

Ângela Freitas, Paula Santana, Cláudia Costa, Artur Vaz
University of Coimbra, Portugal

Population health status evaluation and monitoring is an essential tool for the design of public policies, providing the essential framework to understand and justify choices and supporting the best combination of available resources and expected outcomes. This requires multidisciplinary answers to overcome the traditional and exclusive medical approach of health. Advances in the methods used to measure health status have taken place during the past decade, although more valid, holistic, comprehensive, and standardized, ways of measuring and reporting population health status may be needed to improve opportunities for comparability.
This presentation forms part of the GeoHealthS project which aimed evaluate and monitor Portuguese population health status in the last two decades; it's being developed a Population Health Status Index (INES) that considers multiple dimensions of population health and decision-maker judgements of different knowledge areas. The construction of INES involves a social-technical approach that integrates the technical elements of a multicriteria model and the social elements of a Delphi process. This methodology (index-building process) includes 1) the identification and selection of indicators that influence and characterize population health status, 2) the relative weighting of such factors with impact on population health (social determinants) but also of those indicators or proxies (such as mortality or morbidity rates) sensitive to express health outcomes, taking into account their capability of explain health status, and 3) the assessing of value functions and strategic weights, and (4) the aggregating of value scores that gives overall value of the population health.
As an aggregated social indicator, it will allow the evaluation and monitoring of population health status against each evaluation criterion individually and on each area of concern, giving all the information needed to rank municipalities by population health status and to monitor its progress (from 1991 to 2011). With these outputs we will be able to

address the problems and challenges of each geographic area, which will give the indispensable framework to guide future public policies interventions.
This kind of issues is still more relevant if one considers the present context of economic and financial crisis. Public resources are under additional strain and there is a strong need to adequate and improve decision making processes on policies to reduce or eliminate health differences among citizens.
Key Words: Health Inequalities; Social Determinants; Health Status Index; Multicriteria Decision Model

New results for separating systems

Freixas Josep, Molinero Xavier
Universitat Politècnica de Catalunya, Spain

Some related and open problems related with strict separating systems are considered: the distribution of the tolerance, the maximum tolerance, the G-tolerance and the maximum G-tolerance; to provide tables of all, up to isomorphism, linearly separable switching functions of a reasonable high number of variables with: a strict separating system achieving its tolerance and its G-tolerance; and to develop an efficient algorithm able to calculate the tolerance and the G-tolerance with a reasonable high number of variables.
Key Words: Strict separating systems, tolerance

Using Filtered Historical Simulations in Option Pricing

Bart Frijns[1], Thorsten Lehnert[2], Remco Zwinkels[3]
[1]Auckland University of Technology, New Zealand, [2]Luxembourg School of Finance, Luxembourg, [3]Erasmus School of Economics, The Netherlands

While there is a large literature that looks at the impact of noise trader risk on stock prices, only a few studies investigate how option prices are affected by behavioral biases. In this paper, we extend the pricing framework of Frijns et al. (2010) by introducing a third group of traders that form beliefs about future volatility based on (presumably) noisy signals from aggregated flows in and out-off domestic US equity mutual funds. Our option valuation framework is similar to a stochastic volatility model and is implemented using a modified filtered historical simulation approach.
We assume that traders in the options market can be classified into three different groups that have different expectations about the future evolution of index volatility: fundamentalists, who trade on the principle of

mean reversion; chartists, who trade on exogenous shocks; and noise traders, who trade on noisy signals from the mutual fund market.

Typically, stochastic volatility option pricing models are implanted using Monte Carlo simulation techniques. Barone-Adesi et al. (2008) propose to calibrate the model using the Filtered Historical Simulation approach, by assuming that et is a random variable drawn from the empirical distribution. In order to estimate our more general volatility dynamics including the noise traders, we develop a modified version of this technique. We make use of the fact that we can match the average fund flows – our indicator for noise trader sentiment' – with the standardized residuals for all days that we use to estimate the empirical distribution. Therefore, in the calibration of the models by sampling from the empirical distribution, we obtain not only the historical standardized residual, but pairs of the news innovations and the noise signals. The method has the implicit advantage that we pick up the exact historical correlation between returns and fund flows without the need to explicitly model them as a separate process.

In our empirical application, we calibrate all models on each Wednesday of the years 2005, 2006 and 2007, and calculate the RMSE's in-sample as well for one week out-of-sample.

In line with Barone-Adesi et al. (2008), we find that calibrating the model with a nonparametric innovation distribution compared to standard Monte Carlo techniques results in slightly lower pricing errors. When we incorporate noise trading in the pricing model, we recognize that chartists and noise traders incorporate their information into their beliefs about volatility in a similar way. They expect volatility to go down for good news (positive returns or fund flows) and to go up for bad news (negative returns or fund flows). However, once we include noise traders' trades, in-sample as well as out-of-sample pricing errors of the models are reduced. Given that we only make minor modifications to the specifications, the differences are sizeable. Results further suggest that on average 25% of traders follow a fundamentalist strategy, 19% a chartists strategy and, as a result, about 56% a noise traders strategy. Additionally, noise traders react more strongly to fund flows for shorter maturities compared to longer maturities, but long-term investors are more sensitive to differences in forecasting performance between the chartists and noise trader strategies compared to short-term traders.

Hence, we present an interesting alternative to the well-known Monte Carlo simulation techniques and investigate the benefits in an option pricing exercise. In line with Han (2008), we show that noise trader risk is an important determinant of prices and that noise trading can partly explain the volatility dynamics underlying option prices.

Detecting regime changes in Markov models

Jesus Enrique Garcia, Veronica Andrea Gonzalez-Lopez
IMECC-UNICAMP, Brazil

Let C be a data collection, indexed by time. C={D(t1),...,D(tn)}, where D(ti) was collected at time ti, ti ≤ tj if i ≤ j. Also, each D(ti) follows a Markovian model with finite alphabet A, denoted by M(ti). We device a consistent procedure to detect changes in the model at time ti0 that allows to decide if D(t{i0}) and D(t{i0-1}) are coming from the same Markovian source. The procedure is based on the equivalence relationship introduced by the Minimal Markov Models (see [7]), that allows to associate to each Markovian model a minimal number of parameters enough to describe a Markovian source. The model´s estimation can be consistently performed through the Bayesian Information Criterion, see for details [1] and [7], also for related topics about those models, see [8] and [10] and [3]. Under the possibility of regime change, we can have situations in which D(t1),..., D(t{i0-1}) are coming from a Markovian m! odel, M(t{i0-1}) different to the Markovian model M(t{i0}) appropriated for D(t{i0}). We apply the procedure to detect prosodic changes from classical to modern European Portuguese (see [2], [4], [5]). Taking in consideration that rhythm is a consequence of several characteristics, like number of syllables in the words, position in the word of the stressed syllable, simple and complex syllabic structure, etc., is possible to look for temporal changes in the rhythm, using written texts. In this context, each D(ti) is a written text in European Portuguese and ti is the author's date of birth from 16th century to the 19th century. In this analysis we detect two main change points, the first one at the turn of the 16th century to the 17th century. The second one, in the second half of the 18th century that spreads to the end of the century. Our findings complement the results attained in [4], which study the changes of the European Portuguese in the same period of time, through the analysis of clitic placement. The processing and types of statistical models are completely defined by the differentiated nature of the data. For example, for acoustic signal processing see [6] and for recent research about the statistical modeling see [8], [9] and [10].

Key Words: Minimal Markov models; model selection; Bayesian information criterion; historical linguistics.

References:

[1] Csiszar, I. and Talata, Z. (2006) Context tree estimation for not necessarily finite memory processes, via BIC and MDL. IEEE Trans. Inform. Theory, 52, 1007–1016.

[2] Frota, S., Galves, C., Vigário, M., Gonzalez-Lopez, V. and Abaurre, B. (2012) The phonology of rhythm from Classical to Modern Portuguese. Journal of Historical Linguistics, 2(2), 173-207.

[3]Galves, Antonio, Charlotte Galves, Jesús E. García, Nancy L. Garcia, and Florencia Leonardi (2012). Context tree selection and linguistic rhythm retrieval from written texts. Annals of Applied Statistics, 6(1), 186-209.
[4]Galves, C., Britto, H. and de Sousa, M. C. P. (2005). The Change in Clitic Placement from Classical to Modern European Portuguese. Journal of Portuguese Linguistics, 4(1), 39-67.
[5]Galves, C. and Faria, P. (2010). Tycho Brahe Parsed Corpus of Historical Portuguese. http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html
[6]Garcia, Jesus, Ulrike Gut, and Antonio Galves (2002). Vocale-a semi-automatic annotation tool for prosodic research. In Speech Prosody 2002, International Conference.
[7]Garcia, J. and Veronica A. Gonzalez-Lopez (2010). Minimal markov models. arXiv preprint arXiv:1002.0729.
[8]García, Jesús E., González-López, V. A., Viola, M. L. L. (2012). Robust model selection and the statistical classification of languages. In AIP Conference Proceedings vol. 1490, p.160.
[9]García, Jesús E., González-López, V. A., and Nelsen R. B. (2013). A new index to measure positive dependence in trivariate distributions. Journal of Multivariate Analysis. 115, 481-495.
[10]García, Jesús E., González-López, V. A., Viola, M. L. L. Robust Model Selection for Stochastic Processes. Communications in Statistics-Theory and Methods. (Forthcoming)

Estimation of the structure of interacting coordinates for a multivariate stochastic process

Jesus Enrique Garcia, Veronica Andrea Gonzalez-Lopez
IMECC-UNICAMP, Brazil

In this paper, we address the problem of estimating the interaction structure between the coordinates of a multivariate Markov chain. In our setting, we have a set of k different sources. At each time t, each source produces a letter in the alphabet A={0,1}. The sources interact between them depending on the past states of the set of k sources.
We introduce a new methodology which estimate a partition of the past such that two possible pasts are in the same part of the partition if and only if, the set of sources interacting are the same when conditioned to each of the two pasts. In addition, we obtain for each possible past, the set of interacting sources.
We can imagine a very simplified model of interacting neurons. Consider a set of k neurons. Discretize time in intervals of a fixed size. For each time interval we will say that the value of a particular neuron is one if there was (at least) one spike from that neuron in that interval of time.

In this setting, we say that two neurons interact if the fact that one of them fire, change the probabilities of the other one to fire. This interaction can be in any of two kinds; it can increase or decrease the probability of firing for the other neuron.
In this work we will suppose that the interaction between the neurons does not necessarily have the same dependence from the past that the joint probability of the set of k neurons. In this way, we admit the possibility of having different structures of dependence from the past for the marginal probabilities and for the interaction between the sources. In other words we allow the possibility of having different marginal distributions with the same set of neurons interacting. All along this work we will use the family of the partition Markov models (see [3]) which is a generalization of the variable length Markov chain family (see [1], [2], [4] and [5], [6] and [7]).
Key Words: Multivariate stochastic process; Partition Markov models; interaction model selection; interacting neurons model.
References:
[1] Csiszar, I. and Talata, Z., "Context tree estimation for not necessarily finite memory processes, via BIC and MDL", IEEE Trans. Inform. Theory 52, 1007–1016, 2006.
[2] Galves, A., Galves, C., Garcia, Jesus. E., Garcia, N. L. and Leonardi, F., "Context tree selection and linguis- tic rhythm retrieval from written texts", Annals of Ap- plied Statistics, 6 1, 186 – 209, 2012.
[3] Garcia, J. E. and Gonzalez-Lopez, V. A., "Minimal Markov Models". arXiv preprint arXiv:1002.0729, 2010.
[4] Garcia, J. E., Gonzalez-Lopez, V. A. and Viola, M. L. L., "Robust model selection and the statistical classi- fication of languages", AIP Conference Proceedings, vol. 1490, p.160, 2012.
[5] Rissanen J., "A universal data compression system", IEEE Trans. Inform. Theory 29(5), 656 – 664, 1983.
[6] Viola, M., L. (2011). Topicos em selecao de modelos markovianos. PhD Tesis. http: //www.bibliotecadigital.unicamp. br/document/?code=000844289
[7] Weinberger, M., Rissanen, J. and Feder, M. (1995). A universal finite memory source, IEEE Trans. Inform. Theory 41(3) 643 – 652.

Graphical modelling of high dimension processes: An environmental application

Ali S. Gargoum
Dept. of Statistics, CBE, UAE University, United Arab Emirates

Graphical modelling plays an important role in providing efficient probability calculations in high dimensional problems (computational efficiency). This issue was discussed in a number of papers by Kim and

Pearl (1983), Pearl (1986), Lauritzen and Spiegelhalter (1988) and Spiegelhalter et al. (1993).
In this paper we discuss one of such problems, namely, we describe fragmenting puff models and some distributional assumptions concerning models for the instantaneous, emission readings and for the fragmenting process in case of contamination dispersal. A graphical representation in terms of a junction tree of the conditional probability breakdown of puffs and puff fragments is proposed.
Key Words: Graphical modeling, Influence diagrams, Junction trees

Respiration Rate Recovery from Accelerometric Data

Franco Garofalo, Francesco Lo Iudice, Giovanni Mancini, Michele Pugliese
Università di Napoli "Federico II", Italy

In this paper, we present a novel method, based on the Hilbert transform, for respiration rate recovery using data from MEMS accelerometers. Monitoring respiration is essential in many medical applications and specific biomedical instrumentation is available to perform the task in ambulatory settings. Nevertheless, very often there is the need to monitor respiration continuously and out of health care centers. In such cases, usually, a trade-off must be found between accuracy and expensiveness of the measurements. In this scenario, MEMS accelerometers have proved to be suitable for the task. Several researchers have faced the issue of extracting respiration rate from accelerometer signal. As a result, two families of techniques have emerged, one collecting methods based on Fourier analysis, and the other one comprising algorithms counting single or multiple threshold crossing. Both of them, though, present some pitfalls. Namely, methods based on Fourier transform implicitly assume the signal to be stationary, or at least to present a dominant frequency. Nevertheless, as the respiratory signal is well-known to be describable as a sinusoid with time-varying amplitude and frequency, i.e. a signal r(t) such that
$r(t) = A(t)\sin(\omega(t)t+\varphi)$
it is clear that, especially for irregular breathing, such assumptions do not hold. On the other hand, while algorithms counting single or multiple threshold crossings cope perfectly with variable frequencies, they still perform weakly when large amplitude variations occur, which is the case especially with accelerometric data. Furthermore, as the respiration signal acquired from accelerometric data is heavily stochastic due to noise and artifacts, multiple local maxima and minima may arise which can drive in error threshold methods. In such case, algorithmic methods find a totally arbitrary trade-off between the chance of overestimating respiration rate due to such local extrema, and the risk of

underestimating it due to some breaths that do not qualify as such for their amplitude is not sufficient.
In attempting to overcome such drawbacks, we introduce a novel respiration rate recovery method based on the Hilbert transform. To this aim, in our paper, we give some mathematical preliminaries and illustrate the proposed method attempting to theoretically explain how it compares with the existing techniques. The theoretical analysis is complemented with extensive experimental validation of the proposed method that refers to experiments in which a tri-axial accelerometer has been fixed to an abdominal band. We have performed extensive measurements in a variety of settings in order to test our method in different conditions of noise and artifacts.
Key Words: sensory systems, Hilbert transform, respiration rate recovery

A spatial regression framework for small area population projection with census data

Ghislain Geniaux, José Lopez Kolkovsky
INRA SAD Ecodéveloppement UR 767, France

Projecting population for small areas are important to help planners and local government in designing their future plans and territorial policies, and constitute a key driver in land use change simulation models. However, despite a long history (Lee and Goldsmith, 1982), the literature on small area population projection stay relatively sparse (Kanaroglou and al. 2008). The objective of this paper is to develop and implement a small area direct population forecasting model relying on a regression approach with spatial dependence. We use mortality, fertility and migration data of French census data at municipal scale between 1962 and 2009. Rich information at the municipal scale on migration allows us to include in the modeling framework the main phenomena and drivers that play a role in sub-regional migration, like socioeconomic and geographic variables. Total munic- ipal population projection estimates at mid! dle term horizon (ten years between census in France) will be included (in a second stage, not discussed here) among the drivers of total number of houses model (Urbanismul Project). We show, start- ing with a very simple multi-regional flow models, that a direct population change model with detailed information on migration at small scales leads naturally to spatial econometric models in which spatial autocorrelation is used both for deal- ing with the multiple spatial scales of the studied processes (fertility and mortality versus migration) and with the presence of spatial autocorrelation in disturbance and/or endogenous variable. It allows to propose simpler models than Kanaroglou and al. (2008) with an explicit integration of spatial

dependance in the migration process. The proposed regression framework appear easily usable in conjunction with simple or cohort components models (Booth 2006, Wilson 2011). However reduced form approach produces more accurate forecast whe! n the interest variable is the total population, because. We p! ropose several types of spatio-temporal regression model to deal with various forms of serial correlation, spatial autocorrelation and unobserved spatial heterogeneity: time autoregressive distributed lag (ARDL) combine with spatial dependance in disturbance term (Spatial Error Model -SEM and Spatial Durbin Model- SDM, Anselin 1988), in endogenous variable (Spatial Lag Model -SLM) or both (SARAR, Anselin 1988). We also propose mixed geo- graphically weighted regression (Geniaux et al 2011) to deal with spatially varying coefficient models. We compare forecast accuracy of spatial regression model to simple extrapolation method (linear extrapolation) and simple/cohort components methods. The regression approaches proposed outperform the other methods in terms of Mean Algebraic Percentage Error (MALPE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Percentage Error (RMSPE) for french census data.

Key Words: Small area population forecast, multi-regional migration model, spatial econo- metrics spatial autocorrelation, spatially varying coefficient models..

References

- Booth, H. 2006. “Demographic forecasting: 1980 to 2005 in review”, International Journal of Forecasting, 22 (3), 547–581.
-Chi, G. and Voss, P. R. (2011). Small-area population forecasting: borrowing strength across space and time. Population, Space and Place, 17(5):505–520.
- Geniaux, G, J.-S. Ay, and C. Napoléone. 2011. “A spatial hedonic approach on land use change anticipations”, Journal of Regional Science, 51 (5), 967–986. URL http://dx.doi.org/10.1111/j.1467-9787.2011.00721.x
- Kanaroglou, P. S, H. F. Maoh, B. Newbold, D. M. Scott, and A. Paez. 2009. “A demographic model for small area population projections: an application to the census metropolitan area of hamilton in ontario, canada”, Environment and planning. A, 41 (4), 964.
- Lee, E. S and H. F. Goldsmith. 1982. Population estimates: methods for small area analysis, ., Sage Publications Beverly Hills, CA.
- Wilson, T. December 2011. “A review of sub-regional population projection methods”, university of Qieensland working paper

Time Operator and Innovation. Applications to Financial Data

Gialampoukidis Ilias, Antoniou Ioannis
Aristotle University of Thessaloniki, Greece

The theory of Time Operators has recently been applied into real life problems with the estimation of the innovation probabilities. Based on the assumption that the asset values follow Geometric Brownian Motion with constant variance within each trading day, the internal Age of an asset turns out to be a new statistical index, assessing the average innovations. Moreover, the unpredictability of the t-th observation Xt is estimated by the distribution of innovations of X_t. The innovation probabilities and internal Age are estimated using nonlinear stochastic volatility models.
Key Words: Time Operator, Innovation, Financial Data, Stochastic Volatility models.

Functional clustering by smoothing quantile regression

Paolo Girardi, Roberto Pastres, Carlo Gaetan
Department of Environmental Sciences, Informatic and Statistic; University Ca' Foscari of Venice, Italy

This work presents a functional clustering procedure applied to environmental time series of a physical parameter (the chlorophyll type A concentration) in the coastal area of the Adriatic Sea. The data for the classification analysis is formed by glob-colours data during the period 2002–2012 (monthly values, 11 calendar years) provided by the ACRI server (http://hermes.acri.fr/) using satellite data source combining information of MERIS, Seaways and MODIS optical sensors. The choice of a basis implies the type of features of the series that are to be enhanced or hidden in the representation. Our proposal combines time series interpolation with smoothing quantile splines and the agglomerative clustering algorithm, such as partitioning around medoids technique. Our final purpose is to obtain a classification of the coastal areas in to homogeneous zones in order to select areas at high impact of chlorophyll type A concentrations. The analysis was performed by R software. This approach permits to take into account the quantile of interest and to calculate a more robust clustering procedure respect to other classical methods.

Analysis of information into marginal effects

Valérie Girardin
Laboratoire de Mathématiques N. Oresme, Université de Caen Basse Normandie, France

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. The log-linear parametric model is obtained by minimizing the Kullback-Leibler divergence relative to the reference probability under information constraints.
The model was introduced by J. Zighera (1985) for studying the work of women in Europe and formalized by V. Girardin and A. Ricordeau (1999). By construction, its parameters satisfy constraints that lead to an additive analysis of information into marginal contributions -- thus bear a meaning in terms of information.
Via a parametrization of the model which ensures its identifiability, we determine the asymptotic distribution of the involved statistics, with explicit variance. This allows us to select a suitable model through testing the influence of each marginal effect.
The present contribution to the model is applied to PISA data (Programme for International Student Assessment of the OECD), for studying the influence of time and personnal background on the observed performance inequalities between school students in France.
Joint work with Justine Lequesne (UCBN) and Olivier Thévenon (INED, Paris)

Keywords: information, contingency tables, Kullback-Leibler divergence, log-linear model

References:

1. Girardin, V., Ricordeau, A., (1999), Analysis of information into margins : a log-linear parametric approach, Rapport technique, 1999/14, Université de Caen.
2. Thévenon O. (2008) Labour force participation of women with children: disparities and developments in Europe since the 1990s. Economix Université Paris X
3. Zighera J. (1985) Partitioning information in a multidimensional contingency table and centring of loglinear parameters App. Stoch. Mod. Data An., V1, p93-108

A New Value-at-Risk Semi-parametric Estimation Procedure

M. Ivette Gomes[1], Frederico Caeiro[2], Fernanda Figueiredo[3]
[1]Universidade de Lisboa, FCUL, DEIO and CEAUL, [2]Universidade Nova de Lisboa, FCT and CMA, [3]CEAUL, and Universidade do Porto, Faculdade de Economia, Portugal

For any level *p*, $0 < p < 1$, and on the basis of a sample $\underline{X}_n := (X_1, \mathrm{K}, X_n)$ of either independent, identically distributed or possibly weakly dependent and stationary random variables from an unknown model F with a heavy right-tail function, the value-at-risk (VaR_p) at the level p, i.e. the size of the loss occurred with small probability *p*, is estimated by a new semi-parametric procedure. With the notation $(X_{1:n} \le \mathrm{L} \le X_{n:n})$ for the set of ascending order statistics associated with the available sample $\underline{X}_n$, such a procedure is based on $A_r(k)$ the mean-of-order-*r* of the set of statistics

$$\underline{U} := \left\{ U_{ik} = X_{n-i+1:n} / X_{n-k:n}, 1 \le i \le k < n \right\}$$

with *r* a real number and *k* the number of top order statistics used in the estimation. On the basis of $A_r(k)$ it is possible to build a consistent estimator of the extreme value index (EVI), ξ, denoted $H_r(k)$, studied for $r \ge 0$ in Brilhante *et al.* (2013), and generalizing the classical Hill (Hill, 1975) EVI-estimator, $H_0(k)$. It is next possible to build a reduced-bias EVI-estimator, $\bar{H}_r(k)$. The new class of VaR_p-estimators is similar to the Weissman-Hill VaR_p-estimator studied in Weissman (1978), but with a functional expression given by

$$VaR_p(k) := X_{n-k:n} (k/(np))^{\bar{H}_r(k)} .$$

An algorithm for the choice of the tuning parameters (*k, r*) is suggested and applied to both real and simulated data sets.

Keywords: Heavy right-tails, Monte-Carlo simulations, Statistics of univariate extremes, Value-at-risk estimation

References

1. Brilhante, M.F., Gomes, M.I. and Pestana, D., A simple generalization of the Hill estimator, Comput. Statist. and Data Analysis, 57:1, 518–535 (2013).
2. Hill, B.M., A simple general approach to inference about the tail of a distribution, Ann. Statist., 3, 1163–1174 (1975).
3. Weissman, I., Estimation of parameters and large quantiles based on the k largest observations, J. Amer. Statist. Assoc., 73, 812–815 (1978).

On the probabilistic structure of power TGARCH models and applications to real data

Esmeralda Gonçalves[1], Joana Leite[2], Nazaré Mendes-Lopes[1]
[1]CMUC, Department of Mathematics, University of Coimbra, [2]CMUC, Instituto Politécnico de Coimbra, ISCAC, Portugal

Conditionally heteroscedastic models appeared in literature, introduced by Engle [2], with the aim of describing features of non-linear behavior in time series for which, in particular, the conditional volatility is strongly dependent of past observations. The asymmetrical reaction of the volatility according to the sign of past observations led to threshold ARCH models in which the conditional standard deviation of the process at time t is a piecewise linear function of negative and positive values of past observations. Similarly, the presence of long memory in the shocks of the conditional variance contributed to the proposal of power conditional heteroscedastic models, by Ding et al. [1] among others. Following this idea, we developed in Gonçalves et al. [3] a natural extension of TGARCH processes that allows to take into account both long memory and asymmetry in the stochastic volatility, namely the δ-TGARCH model with real power δ and general error process.

The probabilistic structure of this model is analyzed and we establish conditions of stationarity, ergodicity and δ-order moments existence and also a representation of the conditional volatility as function of the present and past observations. We apply these processes to the study of the dynamical behavior of daily solar activity, based on the plage region areas observed in each one of the North and South solar hemispheres and measured in the Ca II K3 Coimbra's spectroheliograms between 1976 and 2006, Gonçalves et al [4]. Our study indicates that the temporal evolution of these series is well described by ARMA processes with δ-TGARCH errors and that the conditional volatility, strongly present in solar activity, is not well reproduced by these modelations with integer power.

Keywords: Stochastic modeling, Time series, Power TGARCH models

References

1. Ding, Z., Granger, C.W., Engle, R.F., A long memory property of stock market returns and a new model, J. Empirical Finance, 1, 83-106, (1993).

2. Engle, R. F., Autoregressive conditional heteroskedasticity with estimates of the variance of the UK inflation, Econometrica, 50, 987-1008 (1982).

3. Gonçalves, E., Leite, J., Mendes-Lopes, N., On the probabilistic structure of power threshold generalized ARCH stochastic processes, Stat. Prob. Lett., 82, 1597-1609 (2012).

4. Gonçalves, E., Mendes-Lopes, N., Dorotovic, I., Fernandes, J.M., Garcia, A., North and South Hemispheric Solar Activity for Cycles 21-23:

Asymmetry and Conditional Volatility of Plage Region Areas, Solar Physics, 289, 6, 2283-2296 (2014).

Modelling errors in temperature forecasts

Rui Gonçalves
LIAAD-INESC TEC and Dept. of Civil Engineering, University of Porto, Portugal

A random variable with a symmetry coefficient and kurtosis coefficient close to 0 and 3, respectively, is called nearly Gaussian. In order to find out if a random variable is nearly Gaussian we start first to test for Gaussianity. For that purpose one can use several goodness-of-fit tests. For the modelling of nearly Gaussian variables one should consider other symmetrical distributions or, by considering a simple data transformation. If a set of data is nearly Gaussian then, a power transformation $Y=X^c$, where $c=(2k+1)/(2j+1)$, with c close to 1, can transform the data to Gaussianity. This transformation should be bijective and is related to the well-known Box-Cox transformation. Hence, if the transformed data Y is approximately Gaussian then $X=Y^{(1/c)}$ will be the power transformation of a Gaussian variable. We can obtain a mathematical formula for both first order ordinary moment and Kurtosis of ! a zero mean Gaussian variable. This property of the Gaussian distribution allows relating the Kurtosis coefficient to the value of the exponent $1/c$. Therefore, for each nearly Gaussian data set, it is possible to identify an exponent that transforms it to Gaussianity. We apply this method to model the errors of predicted maximum and minimum temperatures in the city of Porto on the year 2011.
Key Words: Nearly Gaussian; temperature forecasts; Kurtosis, power transformation

Sudden failure of the neuronal network and its decomposition model

S. Govor
Bauman Moscow State Technical University, Russia

Consider the principles of the model of sudden catastrophic and partial failure of neural network structures, based on the concept of axons in parallel dendrite I/O, consisting of a set of logical elements.
Keywords: Neuronal networks, probability, temporary failures, dendrites

Quality control of GNSS-Receivers by accuracy-based analysis

Federico Grasso Toro[1], Michal Hodoň[2], Jana Púchyová[2]
[1]Institute for Traffic Safety and Automation Engineering, Technische Universität Braunschweig, Germany, [2]Faculty of Management Science and Informatics, University of Žilina, Slovakia

In this paper we present a quality control method for Global Navigation Satellite System (GNSS) receivers.
A Statistical quality control (SQC) approach is proposed, focused on quantitative trueness, precision and position availability analysis of GNSS receivers'. The position availability is described as the percentage of the total received data that can be considered accurate under n-Sigmas boundaries; being n the level of requested accuracy.
As part of this accuracy-based position availability analysis several filters techniques are tested, in order to select the most reliable for this specific quality control method. Traditional statistical methods are compared with Particle Filter (PF) approach and Mahalanobis Ellipses Filter (MEF) approach.
The quality control method is depicted in graphical representation. And the results are analysed from an end-user point of view. A detailed description of the receiver's characteristics and conditions of the measurements are presented as part of a case study. Significant differences between proposed approaches are shown.
Keywords: Quality control, GNSS receivers, Mahalanobis Ellipses Filter, Particle Filter.

Employment and Fertility – A Comparison of the Family Survey 2000 and the Pairfam Panel

Andreas Groll, Jasmin Abedieh
Department of Mathematics, Ludwig-Maximilians-University Munich, Germany

The major objective of this work is the analysis of the relationship of employment and fertility, also regarding causality. First, the basic sociological theories concerning fertility from an employed women's perspective are shortly summarized. Here, the main focus is on the economical approach with regard to fertility. Based on Germany's current panel analysis of intimate relationships and family dynamics (pairfam), Cox's proportional hazards model [1] is used to test the general assumptions concerning fertility and employed women and to validate the study from [2]. There, the effect of employment on the fertility is analyzed for women based on the data of the German Family Survey 2000, using a proportional hazards model with a piecewise constant baseline hazard. They show that in general employed women have a

lower transition rate into motherhood than unemployed women, but corresponding causality tests indicate that this effect is causal only to some extent due to self-selection and anticipated fertility decisions.
On the one hand, we find that the estimated effects for the Cox model based on the pairfam data are surprisingly consistent with the results from [2]. On the other hand, a corresponding indirect causality test cannot confirm the opposite direction, namely that self-selection in terms of anticipated fertility decisions also affects employment. We conclude that with respect to causality a more sophisticated analysis seems necessary.
Keywords: Pairfam, Employment, Fertility, Event data analysis, Cox's proportional hazards model.

References

1. Cox D. R. (1972). Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society B 34, 187–220.
2. Schröder J. and Brüderl J. (2008). Female Labor Force Participation and Fertility: An Analysis of the Transition to Parenthood in West Germany. Zeitschrift für Soziologie 37:2, 117–136.

Estimation of the Present Values of Life Annuities for the Complex Actuarial Models *

Oxana V. Gubina, Gennady M. Koshkin
Tomsk State University, Russia

The paper deals with a problem of estimating the average present values of the life annuities. On the basis of individuals death moments random samples, both parametric and nonparametric estimators are constructed for the models of the whole, n-year term, and m-year deferred life insurance. The asymptotic normality and mean square convergence of the proposed estimators are proved. By simulations, the nonparametric estimates show adaptability, when the distribution is changed, and exceed parametric estimates, oriented on the best results only for the given distributions.
Key words: Parametric and nonparametric estimators, Life annuities, Asymptotic normality, Mean square convergence

*Supported by Russian Foundation for Basic Research, projects no. 13-08-00744, and 13-08-98027

Asymptotic analyses of a cold-stand-by system with large number of units and repair

Guldal Guleryuz, Ebru Yuksel
Hacettepe University, Department of Industrial Engineering, Turkey

In this paper, a system consisting n identical, repairable units with exponential failure rates, when n is very large is considered. There are r repairmen, each working with large repair rate. Operational units can be in working state, or kept as spares. All spare units are assumed to be operational, cold-stand-by units. If a working unit fails and there are spares available, one unit from spares is sent to working state and the failed unit is sent to repair. A repaired unit is assumed to be as good as new. If there are enough working units, a repaired unit is sent back to spares. If not, they are sent to working state. The asymptotic behavior of the trajectory function is investigated under different conditions regarding failure/repair rates, number of repairmen and the number of cold stand by units in the system. The system is simulated for different sets of parameters and different but related models are analyzed in transient phase.
Key Words: Asymptotic analysis, maintenance, simulation

Fractional Difference ARFIMA Models for long memory timeseries

Maryam Haghiri Limoudehi, Dariush Mohamadi
Statistical Centre of Iran, Iran

A class of general models for long memory time series is the fractional differenced models, ARFIMA (p,d,q).This class is a generalization of famous Box – Jenkins ARIMA models, where the parameter d is a real number. The series is stationary and invertible if $-0.5 < d < +0.5$ .These models are decreasing hyperbolically which is more slower than exponential decay for ARMA. When $0 < d < 0.5$, The series have a long memory, and when $-0.5 < d < 0$ they are unstable or antipersistent.
In this paper, the long memory time series are presented and by defining the predictable memory, we show the methods for choosing the parameters of ARMA adjustment for (0,d,0) along with minimization of prediction variance for one safer ahead forecast.
Key words: Autoregressive processes, Moving average processes, long memory time series, fractional difference methods.

Predicting achievement of first year students: a new approach

Peter G.M. van der Heijden[1], Dave Hessen[2], Theo Wubbels[3]
[1]Utrecht University, Department of Methodology and Statistics, The Netherlands and University of Southampton, UK, [2]Utrecht University, Department of Methodology and Statistics, Utrecht, the Netherlands, [3]Utrecht University, Department of Education, The Netherlands

In this paper a new approach to study achievement is presented. The response variable 'Achievement' is coded as an ordinal variable with levels (i) Dropout in year 1, (ii) Dropout after year 1 or studying longer than 4 years, (iii) Diploma in year 4, and (iv) Diploma in year 3. Achievement is predicted using ordinal regression, using data available in the student register of the Faculty of Social and Behavioral Sciences of Utrecht University. In the presentation of the results, an interpretation in terms of probabilities for the students to end in one of the four levels of the variable 'Achievement', is emphasized. Analyses are carried out at different points in time, where it is found that exam results become increasingly important. An Excel programme is presented, that, for a student, assesses his or her probabilities for the 'Achievement' categories after filling in his or her characteristics. The methodology presented can be easily implemented by other institutions in higher education.
Keywords: Prediction; logistic regression; ordinal regression

Performance analysis of finite source, non-preemptive priority queueing systems

Sedda HAKMI, Ouiza LEKADIR, Djamil AÏSSANI
University of Béjaia, Algeria

This work is devoted to the modeling and performance evaluation of non preemptive priority queueing systems with finite source via the Generalized Stochastic Petri Nets (GSPN). Indeed, their extension to Markov processes increases their interest greatly. This allows us to perform qualitative and quantitative analysis of these priority systems. Moreover, we show how this high level formalism allows us to cope with the complexity of this finite source, priority systems.
Key Words: Queueing systems, Non-preemptive priority, Performance evaluation, Generalized stochastic Petri nets.

Modeling a Free-Ranging Rhesus Macaque Population

Raisa Hernandez-Pacheco
University of Puerto Rico, Puerto Rico

Modeling free-ranging animal populations is often hindered by hidden mechanism of population regulation. The lack of long term demographic data has hampered the parameterization of density-dependent models for the study of population dynamics, as well as viability analyses for management or conservation plans. In this study I make use of the most extensive nonhuman primate demographic data available - the Cayo Santiago's rhesus macaque colony, the oldest free-ranging monkey colony in the world. The colony is monitored and managed for population control using harvesting strategies. A density-independent and a density-dependent population matrix model were parameterized and their projections compared with the observed population changes over three decades. I also carried out sensitivity analyses and applied them to each projection model in order to determine the optimum harvesting regime that would ensure population viability. The population exhibited negative density-dependence in fertility and the model incorporating this relationship accounted for 98% of the observed population dynamics. Variation in survival and fertility of sexually active individuals contributed the most to the variation in long-term fitness. Therefore, targeting culling to adult individuals would contribute significantly to population control. These findings are novel in describing density-dependent dynamics in a provisioned primate population and in using it for management purposes. Moreover, density-dependent is a timely ecological problem as it may become stronger in wild populations due to increasing habitat loss and food scarcity. This study demonstrates the importance of incorporating variation in population size into population projection models for a better understanding of the mechanisms regulating population growth and, therefore, its viability.
Key Words: Density-dependence, Matrix projection models, Population projection, Rhesus macaques, Sensitivity analysis,

Bayesian Nonparametric Estimation in Nonlinear Dynamic Systems with Geometric Stick Breaking Random noise

Spyridon J. Hatjispyros, Christos Merkatas
University of the Aegean, Department of Mathematics (Statistics and Actuarial-Financial Mathematics), Greece

In this paper we propose a Bayesian Nonparametric Mixture model for the full reconstruction, from observed time series data, of polynomial

discrete time random dynamical systems, based on the Geometric Stick Breaking (GSB) process introduced by Fuentes-García et al. (2010). We provide a comparison study with models using Dirichlet process based mixtures. We demonstrate the inference procedures when the functional form of the reconstruction equation is cubic and the nonparametric component is applied to additive errors. Finally, we claim that for dynamical reconstruction purposes, GSB process priors are sufficient.
Key Words: Geometric Weights, Random Dynamical Systems, Stick Breaking

A survey on the isometry of certain systems orthogonal polynomials in martingale spaces

Edmundo José Huertas Cejudo
Departamento de Matemática, Universidade de Coimbra, Portugal

Sets of orthogonal martingales are important because they can be used as stochastic integrators in a kind of chaotic representation property. In this talk, we revisited the problem studied by W. Schoutens in [1], investigating how an inner product derived from an Uvarov transformation of the Laguerre weight function is used in the orthogonalization procedure of a sequence of martingales related to a certain Lévy process, called Teugels Martingales. Since the Uvarov transformation depends by a $c<0$, we are investigating how to provide infinite sets of strongly orthogonal martingales, each one for every c in $(-\infty, 0)$. In a similar fashion of [2] (see also [3]), our aim is to introduce a suitable isometry between the space of polynomials and the space of linear combinations of Teugels martingales as well as the general orthogonalization procedure.
Key Words: Lévy processes; Stochastic processes; Orthogonal polynomials; Laguerre-type polynomials; Krall-Laguerre polynomials; Inner products;

References

[1] D. Nualart and W. Schoutens, Chaotic and predictable representations for Lévy processes, Stochastic Process. Appl. 90 (2000) 109–122.
[2] W. Schoutens, An application in stochastics of the Laguerre-type polynomials, J. Comput. Appl. Math. 133 (1-2) (2001), 593-600.
[3] W. Schoutens, Stochastic processes and orthogonal polynomials, in Lecture Notes in Statist. 146, Springer-Verlag, New York, (2000)

Understanding the interactions between global and regional seasonality of crude oil consumption

Julian Inchauspe, Jun Li
DMEE, Curtin University, Australia

Seasonality is an important characteristic of world oil consumption and has implications for pricing, policy and investment decisions. It is worth noting that as seasonal patterns some countries have seasonal patterns that exacerbate the seasonal pattern of world consumption (“coincident seasonality”), other countries may have “counter-directional” seasonal patterns that tend to smooth the seasonality at global level. To the best of our knowledge, these observations have not been deeply studied in the literature and give room to some research questions. If consumption in countries such as China continues to grow at a rapid rate, will this exacerbate the seasonality of world consumption? Are there countries which would smooth world seasonality as they grow? Which countries are responsible for seasonality in consumption and what will happen if expected oil consumption for the future materialise?

This paper contributes to the stochastic modelling of commodity inventory literature by consistently modelling the potential seasonality of primary energy demand and supply across OECD and emerging countries with empirically observed properties. We use a comprehensive crude oil dataset which includes major regions over a 20-year period to investigate the trend and stochastic characteristics of variation in energy supply and demand on quarterly. This has extremely important implications for global energy markets design and coordination as well as optimisation of investment in long term energy supply infrastructures.

In this paper, seasonality is modelled as the variation of quarterly crude oil consumption around a long-term trend. Across these regions, these seasonal patterns vary according to climate differences, intra-year variation in agricultural activity, income seasonality and other idysioncratic factors. If quarterly seasonality can be modelled as a predictable, recurring pattern over for each year, it is possible to use this information to make forecasts that apply to upcoming trends in crude oil demand. Capturing seasonality of oil imports is important for both importing and exporting countries, for reasons pertaining to inventory control and production planning respectively.

Time-series econometric techniques are used for modelling seasonal patterns around a long-term trend. Our approach is based on classic demand theory: we assume that quantity crude oil demanded in each country depends on a reference crude oil price, income, price of substitute fuels and seasonal components. Two major techniques are currently being employed in this paper: (i). Individual ADL (autoregressive distributed lag models which include exogenous variables) and (ii). A global VAR (vector autoregressive) model.

The information that is obtained through this process is useful for analytical purposes. First, appropriately accounting for seasonality allows countries and regions to mitigate the risks of unexpected large fluctuations in commodity market. Energy is one of major input factor of economic activities and energy demand is also characterised by inherent seasonal variation due to the heterogeneity of consumers. Second, capturing seasonality of the oil demand is also important for both producers and intermediary traders, for reasons pertaining to inventory control and production planning respectively. Third, it is argued that using quantity –based energy consumption data can allow energy modellers and analysts to identify major seasonal factors and mend the shortcoming of data unavailability of energy prices data.
Our results indicate that most regions exhibit significant demand seasonality over the analysed period with the two models employed so far. The demand variations do not occur simultaneously and showed considerable heterogeneity. However, further improvements will have to be made before reaching definitive conclusions. First, the authors will have to re-consider the advantages and disadvantages of the ADL and global VAR methodologies and choose the most appropriate. Second, it has to be ensured that the long- and short-term characteristics of income are properly modelled. Third, the authors are trying to gather a more complete or comprehensive dataset of price of substitutes (natural gas and coal) across counties. A final section of the paper will include forecasts in which seasonal components are applied.
Key Words: seasonality, stochastic modelling of oil consumption, ADL, VAR.

Moments of non-homogeneous semi-Markov flow

Ivnitskiy V.A.
Moscow State University of Railway Engineering, Russia

There are obtained initial moments of the process $\xi(t)$, meaning a time interval to the next time arrival, and of the process $\nu(t)$, meaning a number of arrivals before a time t, for non-homogeneous semi-Markov flow under general initial conditions in terms of Laplace transformations. Let X - final or accounting multitude with elements $i_1, i_2, \ldots$ It is had non-homogeneous Markov chain $\{\nu_n, n \geq 1\}$ with values to X and transition matrix $((p_{ij}^{(d)}))$ on d step of Markov chain, d =1,2,…., $\sum_{j \in X} p_{ij}^{(d)}$ = 1, $i \in X$. Non-homo-geneous semi - Markov process is random process $\nu_1(t), t \geq 0,$ with properties: 1) to half –interval

$[t_i, t_{i+1})$: $\nu_1(t) = \nu_{i+1}, i = 0,1,...,t_0 = 0,$ 2) under fixed realization Markov chain $\nu_n = i_n,\ n \geq 1,$ sojourn values $t_1, t_2 - t_1,...$ process $\nu_1(t)$ to states $i_1, i_2,...$ positive and independent, sojourn value $z_{\nu_n,\nu_{n+1}}^{(n+1)}$ on $m+1$ step depend on state ν_n, in which process $\nu_1(t)$ is, and on next state ν_{n+1}; it are had following distribution functions $P\{t_n - t_{n-1} < x / \nu_{n-1} = i,\ \nu_n = j\} = F_{ij}^{(n)}(x), n \geq 1.$ Succession times $t_0, t_1, t_2,....$ make non-homogeneous semi-Markov flow. In every from times t_0, t_1, t_2,.... one customer is arrived.

Find distribution value $\xi(t)$ - residual time from t to next arrival of non-homogeneous semi-Markov flow. We will introduce three – measure random process ($d(t), k(t), \xi_{k(t)}^{d(t)}(t)$) and indicator $I(d(t) = d, k(t) = k,\ \xi_k^{(d)}(t) < x)$, where $d(t)$ - Markov chain step number at time t. Then have $I(\xi(t) < x)$ =

$\sum_{d=1}^{\infty} \sum_{k \in X} I(d(t) = d, k(t) = k,\ \xi_k^{(d)}(t) < x)$. Put $\varphi_k^{(d)}(t,x)$ = $M\ I(d(t) = d, k(t) = k,\ \xi_k^{(d)}(t) < x)$. Let also $\chi_{k_1,k}^{(d)}$ -random value, having meaning 1 with probability $p_{k_1,k}^{(d)}$ and 0 with probability 1- $p_{k_1,k}^{(d)}$, From here $M\ I(\ \chi_{k_1,k}^{(d)}$ =1) = $p_{k_1,k}^{(d)}$.

We use the method of stochastic difference equations (MSDE) [8-11] to obtain the initial moments. One can summarise its essence as follows. It is suggested to construct the difference equations at once for the initial moments of Markov process of desired order first, then to use the operator of the mathematical expectation, then to take the limit and finally obtain a system of differential equations for the moments. This method turns out to be more effective for obtaining the arbitrary order non-stationary moments, but area of it applications is more narrow.

Theorem 1. *For non-stationary state probability* $\varphi_k^{(d)}(t,x)$ *of three – measure random process* ($d(t), k(t), \xi_{k(t)}^{d(t)}(t)$) *following system* of *linear differential equation with private derivatives*

$$\frac{\partial}{\partial t}\varphi_k^{(d)}(t,x) - \frac{\partial}{\partial x}\varphi_k^{(d)}(t,x) = -\frac{\partial}{\partial x}\varphi_k^{(d)}(t,0) + \sum_{k_1 \in X} \frac{\partial}{\partial x}\varphi_{k_1}^{(d-1)}(t,0)\ p_{k_1,k}^{(d)}\ F_{k_1,k}^{(d)}(x), k \in X, \quad (1)$$

holds with initial conditions $\varphi_i(0,x) = \varphi_i^{(0)}(x) \geq 0,\ \varphi_i^{(0)}(\infty) = p_i^{(0)},\ i \in X,\ \sum_{i \in X} p_i^{(0)} = 1,.$

Put $\widetilde{\varphi}_k^{(d)}(u,s) = \int_0^\infty e^{-ut-sx} d_x\, \varphi_k^{(d)}(t,x),\ \frac{\partial}{\partial x}\widetilde{\varphi}_k^{(d)}(u,0)$

$= \int_0^\infty e^{-ut} d\, \frac{\partial}{\partial x}\varphi_k^{(d)}(t,0),\ \widetilde{\varphi}_k^{(0)}(s) = \int_0^\infty e^{-sx} d_x\, \varphi_k^{(0)}(x),\ \widetilde{f}_{k_1,k}^{(d)}(s) =$

$\int_0^\infty e^{-sx} d_x\, F_{k_1,k}^{(d)}(x),\ k_1, k \in X.$

Theorem 2. *The following formula holds for* $\widetilde{\varphi}_k^{(d)}(u,s)$

$$\widetilde{\varphi}_k^{(d)}(u,s)\ (u-s) = -\frac{\partial}{\partial x}\widetilde{\varphi}_k^{(d)}(u,0) + \sum_{k_1 \in X} \frac{\partial}{\partial x}\widetilde{\varphi}_{k_1}^{(d)}(u,0)\ p_{k_1,k}^{(d)}\ \widetilde{f}_{k_1,k}^{(d)}(s) + \widetilde{\varphi}_k^{(0)}(s),\ k \in X, \quad (2)$$

where $\frac{\partial}{\partial x}\widetilde{\varphi}_k(u,0),\ k \in X$, *is obtained by the following system of linear algebraic equations*

$$\frac{\partial}{\partial x}\widetilde{\varphi}_k^{(d)}(u,0) = \sum_{k_1 \in X} \frac{\partial}{\partial x}\widetilde{\varphi}_{k_1}^{(d-1)}(u,0)\ p_{k_1,k}^{(d)}\ \widetilde{f}_{k_1,k}^{(d)}(s) + \widetilde{\varphi}_k^{(0)}(u) = 0, \quad k \in X. \quad (3)$$

Let us consider the process $\nu(t)$ which is not markovian. One can extend the process $\nu(t)$ by two supplementary variables: above mentioned $\xi(t)$ and $k(t)$- state Markov chain at time t. Three-measure vector $\zeta(t) = (\nu(t), I(d(t) = d, k(t) = k, \xi_k^{(d)}(t) < x))$ already appear Markov vector random process. Yet for decision putting problem it is also little fitly. For markovization of process $\nu(t)$ we shall use supplementary variables without increase of dimension $\nu(t)$.

We call this method
markovization without increase of dimension (MWID for short).

Essence of this method in the present case consists in the multiplication of the initial process $\nu(t)$ on the indicators $I(d(t) = d, k(t) = k)$ and $I(\xi_k^{(d)}(t) < x)$, that is we use the Markov

processes $\nu_k(t,x) = \nu(t)\, I(\xi_k^{(d)}(t) < x)\, I(d(t) = d, k(t) = k)$, $k \in X$, and apply to it the MSDE. Put $n_k^{(d)}(t,x) = M\, \nu_k^{(d)}(t,x)$, $k \in X$.

Theorem 3. *The following system of linear differential equations with private derivatives for* $n_k^{(d)}(t,x)$, $k \in X$, *holds*

$$\frac{\partial}{\partial t} n_k^{(d)}(t,x) - \frac{\partial}{\partial x} n_k^{(d)}(t,x) = - \frac{\partial}{\partial x} n_k^{(d)}(t,0) +$$

$$\sum_{k_1 \in X} \frac{\partial}{\partial x} n_{k_1}^{(d-1)}(t,0)\, p_{k_1,k}^{(d)}\, F_{dk}(x) +$$

$$\sum_{k_1 \in X} \frac{\partial}{\partial x} \widetilde{\varphi}_{k_1}^{(d-1)}(u,0)\, p_{k_1,k}^{(d)}\, F_{k_1,k}^{(d)}(x),\ k \in X, \tag{4}$$

with initial conditions $n_m(0,x) = n_m^{(0)}(x)$, $n_m(0,\infty) = n_m^{(0)}$, $m \in X$, *where* $\frac{\partial}{\partial x} \widetilde{\varphi}_k^{(d)}(u,0)$, $k \in X$, *is obtained by turning their Laplace transformations, which is obtained by system of linear algebraic equations* (3) *theorem* 2, $\varphi_k^{(0)}(x) = n_k^{(0)}(x) / n_k^{(0)}$, $k \in X$.

Put $\widetilde{n}_k^{(d)}(u,s) = \int_0^\infty \int_0^\infty e^{-ut-sx} d_x\, n_k^{(d)}(t,x) dt$, $\frac{\partial}{\partial x} \widetilde{n}_k^{(d)}(u,0) =$

$\int_0^\infty e^{-ut} \frac{\partial}{\partial x} n_k^{(d)}(t,0) dt$, $\widetilde{n}^{(l)}(u) = \int_0^\infty e^{-ut} \times M\nu(t)^l\, dt$,

$\widetilde{n}_k^{(0)}(s) = \int_0^\infty e^{-sx} dn_k^{(0)}(x)$, $k \in X$, $\widetilde{f}_{k_1,k}^{(d)}(s) = \int_0^\infty e^{-sx} d_x\, F_{k_1,k}^{(d)}(x)$,

$k_1, k \in X$.

Theorem 4. *The Laplace transformation* $\widetilde{n}_k^{(d)}(u,s)$ *is obtained by the formula*

$$\widetilde{n}_k^{(d)}(u,s)\,(u-s) = - \frac{\partial}{\partial x} \widetilde{n}_k^{(d)}(u,0) + \sum_{k_1 \in X} \Big(\frac{\partial}{\partial x} \widetilde{n}_k^{(d-1)}(u,0)$$

$$+ \frac{\partial}{\partial x} \widetilde{\varphi}_{k_1}^{(d-1)}(u,0) \Big)\, p_{k_1,k}^{(d)}\, \widetilde{f}_{k_1,k}^{(d)}(s) + \widetilde{n}_k^{(0)}(s),\ k \in X, \tag{5}$$

where $\frac{\partial}{\partial x}\,\widetilde{n}_k^{(d)}(u,0)$, $k \in X$, *is obtained by system of linear algebraic equations*

$$\frac{\partial}{\partial x}\,\widetilde{n}_k^{(d)}(u,0) =$$

$$\sum_{d\in X} \frac{\partial}{\partial x}\,\widetilde{n}_d(u,0)\;\; p_{dk}\;\widetilde{f}_{k_1,k}^{(d)}(s) + \sum_{k_1\in X} \frac{\partial}{\partial x}\,\widetilde{\varphi}_{k_1}^{(d-1)}(u,0)\; p_{k_1,k}^{(d)}\;\widetilde{f}_{k_1,k}^{(d)}(s) +$$

$$\widetilde{n}_k^{(0)}(u),\; k \in X, \qquad (6)$$

$\frac{\partial}{\partial x}\,\widetilde{\varphi}_{k_1}^{(d-1)}(u,0)$, $k_1 \in X$, *is obtained by system of linear algebraic equations* (3).

*This work is executed with the support of Russian fund fundamental investigations, grant N 13-08-13116

The estimation of seasonal variation of live births in Eastern European countries

Elisabeta Jaba, Christiana Brigitte Balan
Alexandru Ioan Cuza University of Iasi, Romania

Fertility is a demographic phenomenon which plays a key role in the economic development of a country due to its influence on the the population dynamics and to the consequences for the labour force size. Nowadays, the world's population is experiencing a fertility rate that is near or below the replacement level. The evolution of fertility is under the influence of political, social and economic factors.

This paper aims to identify the seasonal patterns in the number of live births' evolution for a sample of Eastern European countries (Poland, Czech Republic, Hungary, Slovakia, and Romania) during the period 1995 – 2012. Before the 90's, these countries experienced demographic policies that were influenced by the communist regimes. Moreover, during the transition to a market economy, they faced new economic constraints that had impact on population's fertility.

The findings underline the specific pattern of the seasonal variations and the common characteristics of the population's fertility behaviour in the observed countries.

Key Words: fertility, live births, seasonal variation, Eastern-Europe

Evolution of electoral behavior by principal axes methods

Margarita Marín Jaramillo, Campo Elias Pardo
Universidad Nacional de Colombia, Colombia

In this paper we study the common voting patterns in Colombian presidential elections between 1986 to 2010. We build contingency tables with subpartitions on rows and columns, where the rows correspond to the Colombian municipalities, according to their population size and the columns correspond to the votes for candidates in each electoral period. We develop a correspondence analysis weighted intra blocks (ACIBP) with cluster analysis to study voting patterns, eliminating the variability induced by population differences and election periods. With this, we conclude that there is an electoral pattern, mainly in the municipalities with population under 50,000, which is mor clear before the 2002 election period.
Keywords: ACIBP, Contingency Tables, Cluster Analysis

Saddlepoint-Based Bootstrap Inference in Dependent Data Settings

Pratheepa Jeganathan[1], Robert L. Paige[2], A. Alexandre Trindade[1], R. Indika P. Wickramasinghe[3]
[1]Texas Tech University, Department of Mathematics & Statistics, [2]Missouri University of Science and Technology, Department of Mathematics & Statistics, [3]Eastern New Mexico University, Department of Mathematical Sciences, U.S.A.

A saddlepoint-based bootstrap (SPBB) method for inference on scalar parameters in the context of time series ARMA models and spatial lattice regression models is presented. The method handles any estimator that is expressible as the root of a quadratic estimating equation (QEE), and includes common estimators like method of moments (MOM), maximum likelihood (ML) and restricted ML (REML). Since the underlying QEE has a moment generating function in closed-form, this is inverted via the saddlepoint method to produce accurate approximations to the respective distributions of the estimators. Confidence intervals are then produced by pivoting the distribution function. The approach relies on the key assumption of monotonicity for the QEE. Alternative strategies for implementing SPBB in the case of non-monotonicity are discussed. SPBB approximations to the underlying densities of the estimators are compared to their corresponding limiting asymptotic counterparts. Simulation studies show that SPBB confidence intervals outperform those based on standard (first-order) asymptotic theory in small to moderate sample size settings, and is orders of magnitude faster than computationally intensive approaches like the bootstrap. The methodology is illustrated on some datasets.

Keywords: Conditional autoregressive model; simultaneous autoregressive model; moving average model; maximum likelihood; quadratic estimating equation; saddlepoint approximation.

Motives, expectations, preparation, and time to degree: A proportional hazards approach with latent covariates

Dimitrios Kalamaras, Aglaia Kalamatianou
Department of Sociology, Panteion University of Social and Political Sciences, Greece

From individual level data for a sample of graduates and students of a Greek university oriented to social and political sciences, we use a Cox Proportional hazards model to investigate the relationship between the risk that a student will either graduate on time or will have a late graduation which may also result in perceptual studentship, and a number of covariates related to students' motives, expectations and preparation. These covariates correspond to directly observable variables as well as to indicator -latent- variables that summarize groups of observable variables, all considered related to time duration of studies. A procedure, mainly used in psychological applications, is implemented as to estimate the scores of the latent variables before put them into the analysis. Gender differences are also examined on the base of the model.

A System for Empirically Effective Credit Risk Analysis

Takeaki Kariya, Yoshiro Yamamura, Zhu Wang
Meiji University, Japan

This paper proposes a statistical system for empirically effective credit risk analysis based on Government Bond (GB)& Corporate Bond (CB) prices. It consists of attribute-dependent GB & CB pricing models. We first define the concept of credit risk price spread (CRIPS)and then Market Rating (MR) method, whose ratings are compared with Agency's ratings in Japanese & US CB market. By useing the MR method, we obtain credit- homogeneous groups and derive term structure of default probabilities for each group.
Key Words: Attribute-dependent bond model, Credit rating, Term Structure of Default Probabilities.

Empirical Credit Risk Analysis on Euro Government Bonds – Term Structures of Default Probabilities–

Takeaki Kariya[1], Yoshiro Yamamura[1], Yoko Tanokura[2], Zhu Wang[3]
[1]GSGB, Meiji University, [2]GSAMS, Meiji University, [3]ZW System, Japan

In this paper, we make a comprehensive credit risk analysis on government bonds (GBs) of Germany, France, Italy, Spain Greece and Germany over the period Arp 2007 to Mar 2012, where interest rate (IR) differential, government bond (GB) price differential, default probability (DP) and CDS are considered. First, applying the GB-pricing model in Kariya et,al (2012) to these GB prices, we derive the term structures of interest rates (TSIRs) and discuss on the Maastricht convergence condition for IR-differentials among these states relative to the German TSIRs. The results are associated with business cycles and budgetary condition of each state.

In the second part, to substantiate this viewpoint, we make credit risk price spread (CRPS) analysis on price differentials and derive the term structures of default probabilities (TSDPs) of the French, Italian, Spanish and Greek GBs relative to the German GBs, where the corporate bond (CB) model proposed in Kariya (2013) is used in the derivation. Then the TSDPs of these GBs are empirically shown to be almost linear functions of the differences of the TSIRs, which enables us to state the Maastricht condition in terms of default probability.

Thirdly the effectiveness of our TSDPs is empirically verified by comparing them with the corresponding credit default swap (CDSs) against US dollars.

Keywords: Government Bond Pricing Model, Credit Risk Price Spread, Term Structures of Interest Rates and Default Probability, Credit Default Swap, Maastricht Treaty

INFLUENCE OF DEMOGRAPHIC FACTORS TO THE AMOUNT AND STRUCTURE OF UNPAID WORK IN SLOVAKIA

ALENA KAŠČÁKOVÁ
Faculty of Economics, Department of Quantitative Methods and Information Systems, Matej Bel University in Banská Bystrica, Slovakia

The article focuses on exploring the demographic factors to the amount and structure of unpaid work of people in Slovak households. Eurostat methodology exactly lists certain activities in a household for which a person does not receive any remuneration. These activities are considered as unpaid work. The data for analysis are drawn from the results of the primary research carried out in 2012 and 2013 years and are compared with the results of the international social research ISSP. Respondents were asked to express how many hours they spent weekly

on average in unpaid work. The data of the both researches are processed from the point of view of demographic factor influences to the total amount of unpaid work in the household and its structure by the individual types of activities.

Comparison of Jump-Diffusion Parameters Using Passage Times Passage

Khaled KHALDI, Khedidja DJEDDOUR, Samia MEDDAHI
Department of Mathematics, Faculty of Sciences, Boumerdes University, Algeria

The main purposes of this paper are two contributions:
(1) it presents a new method, which is the first passage time generalized for all passage times (PT method), in order to estimate the parameters of stochastic Jump-Diffusion process.
(2) it compares in a time series model, share price of gold, the empirical results of the estimation and forecasts obtained with the PT method and those obtained by the moments method applied to the Merton Jump-Diffusion (MJD) model.
Key Words: Merton jump-diffusion, First passage time, Black-Scholes equation, Trajectory, Option pricing

First Passage Time Method Generalization for the Estimation of Stochastic Differential Equation Parameters

Khaled KHALDI, Khedidja DJEDDOUR, Samia MEDDAHI
Department of Mathematics, Faculty of Sciences, Boumerdes University, Algeria

In this paper a new method, generalized passage time, GPT, to estimate the parameters of the Black-Scholes equation is proposed. This approach is based on the first passage time, FPT. We study also the GARCH model and the SDE method for the volatility (SDE-V). Results comparison, parameters estimation of Black-Scholes equation on a time-series model is given.
Key Words: Geometric Brownian, Black-Scholes equation, Trajectory, First passage time, Stochastic volatility

Modeling of spatial redistribution population due to climatic trends

T.R. Kilmatov, O.I. Trinko
Far Eastern Federal University (FEFU), Russia

Spatial heterogeneity of the population density on the planet is expressed especially in East Asia. There is a significant population size

in the south-east and the lowest population in Siberia. Some possible climate predictions involve the warming planet and then the increase of the environment capacity for a favorable living at high latitudes, including Siberia. Some possibility of large-scale migration in the direction the South – the North owing to observable climatic trends is discussed. The mathematical model of interaction of two economic agents considering environment capacity is constructed. Analytical formulas for an assessment of a stream of migration owing to climate change are received. Quantitative estimates of a stream of climatic migration on a labor market including the size and quality of population, labor productivity, salary, volume and cost of the made product, environment capacity for accommodation are presented.

Keywords: climate, production function, environment capacity, labor market, migration

Simultaneous Variable Selection and Outlier Detection for Automated K-means Clustering

Sung-Soo Kim
Korea National Open University, S. Korea

One of the most important problems in cluster analysis is the selection of variables that truly define cluster structure, while eliminating noisy variables that mask such structure. Also outlier detection is a fundamental task while doing cluster analysis. Here we provide an automated process for simultaneous variable selection and outlier identification in doing K-means clustering, which we call automated K-means clustering. The Automated K-means clustering procedure consists of three processes - (i) automatically calculating the cluster number and initial cluster center whenever new variable is added, (ii) identifying outliers for each cluster depending on the used variables, (iii) selecting true variables defining cluster structure in a forward manner. For selecting true variables, we used Brusco and Cradit(2001)'s VS-KM(variable-selection heuristic for K-means clustering) procedure. For identifying outlie! rs we used hybrid approach combining clustering based approach and distance based approach. Simulation result indicates that the proposed automated K-means clustering procedure is very effective at selecting true variables and identifying outliers. Implemented R program can be obtained on the website http://www.knou.ac.kr/~sskim/SimVarAndOutlierKm.R.

Key Words: Automated K-means clustering, Variable Selection, Outlier Detecting, Mojena' stopping rule, VS-KM, Adjusted Rand Index, Mahalanobis Distance

A Diffusion Inventory Model for Perishable Items

Anna Kitaeva, Natalya Stepanova
National Research Tomsk Polytechnic University, Russia

Managing inventory of perishable items has received considerable attention in recent years: see a review by Bakker M., Riezebos J., and Teunter R.H. (2012).

Let the number of independent demands $n >>1$ during a time period T be a stationary process with the expectation $E\{n\} = m_T$ and the variance $Var\{n\} = \sigma_T^2$. For example, for stationary Poisson process $m_T = \sigma_T^2 = \lambda T$.

Demands values are independent identically distributed non-negative random variables with finite first and second moments equal respectively a_1 and a_2.

The items deteriorate continuously with time: at time interval $[t, t + \Delta t]$ an item deteriorates with a probability $p = \kappa \Delta t + o(\Delta t)$ where κ is a deterioration rate coefficient per stocked item. We consider the diffusion approximation of the deterioration process $x(t)$

$$dx(t) = -\kappa x(t)dt + \sqrt{\kappa x(t)}dw_t .$$

The diffusion approximation of the inventory level $Q(t)$ is

$$dQ(t) = -(\kappa Q + m_0)dt + \sqrt{\kappa Q + \sigma_0^2}dw_t ,$$

where $m_0 = \lim_{T\to\infty} \frac{a_1 m_T}{T}$, $\sigma_0^2 = \lim_{T\to\infty} \frac{m_T(a_2 - a_1^2) + \sigma_T^2 a_1^2}{T}$.

We receive the equation for the conditional mean of the remaining lifetime at the beginning of the production cycle given the initial inventory level and find its approximate solution.

The optimal initial inventory level maximizing the profitability is also found.

Keywords: Inventory control, Stochastic model, Perishable items, Diffusion model

References

1. Bakker, M., Riezebos, J., and Teunter, R.H., Review of inventory systems with deterioration since 2001, *European Journal of Operational Research,* **221**, 275–284 (2012).

Entropy Measures and the Generalized Fisher's Information

Christos P. Kitsos, Thomas L. Toulias
Technological Educational Institute of Athens, Greece

This paper investigates the generalized Fisher's entropy type information measure with respect to the multivariate γ–order Normal distribution and

certain boundaries are obtained. Also the Rényi and Shannon entropies are evaluated and discussed.
Keywords: Fisher's entropy type information measure, γ–order Normal distribution, Rényi entropy.

Bandwidth matrix selectors for multivariate kernel regression

Jan Koláček, Ivana Horová
Department of Mathematics and statistics, Faculty of Science, Masaryk University, Czech Republic

The most important factor in multivariate kernel regression is a choice of a bandwidth matrix. This choice is particularly important, because of its role in controlling both the amount and the direction of multivariate smoothing. Considerable attention has been paid to constrained parameterization of the bandwidth matrix such as a diagonal matrix.
Data-driven selector of a full bandwidth matrix for a regression function is considered. The proposed method is based on an optimally balanced relation between the integrated variance and the integrated squared bias. The utility of the method is illustrated through a simulation study and real data applications.
Keywords: multivariate kernel density, unconstrained bandwidth matrix, MSE

Numerical Valuation of American Options with Discrete Dividends

Dessislava Koleva, Mariyan Milev
Sofia University, Bulgaria

We suggest an approach for valuing American options with discrete dividends where the underlying asset follows a geometric Brownian motion. We apply an efficient positivity-preserving method based on a non-standard finite difference scheme in this extreme case where closed-form solutions are still not known. Our method satisfies all requirements of the financial contract such as positivity and smoothness of the numerical solution. The presented approach has a simple programming implementation and the results obtained are well-behaved.
*This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014).
Key Words: American options, discrete dividends, finite difference methods

Robust Normal Two-Armed Bandit and Parallel Data Processing

Alexander V. Kolnogorov
Yaroslav-the-Wise Novgorod State University, Russia

We consider the two-armed bandit problem in application to processing of a large number items of data. Two universal alternative methods of data processing are available, each method has a fixed but a priori unknown probability of successful processing. The core of the problem is that the best method is not known in advance. So, it should be estimated meanwhile the control process.
The usual approach to the control is to process data sequentially, one by one. However, if the problem is considered in minimax setting, it turned out that the control may be implemented in parallel almost without the lack of its quality, i.e. under mild conditions minimax risks in both cases of parallel and sequential controls have close values. For example, T=30000 items of data may be partitioned into N=30 groups each containing K=1000 items of data so that data in each group are processed in parallel and the results of processing are summarized. Calculations show that N=30 provides a high quality of the control.
The usage of summarized results of data processing leads to the following setup of the problem. Distributions of rewards of the two-armed bandit are assumed to be normal with unknown mathematical expectations and unit variances. According to the main theorem of the theory of games minimax strategy and minimax risk are searched as Bayes ones corresponding to the worst prior distribution. In considered case, the worst prior distribution is symmetric and asymptotically uniform. This allows to use numerical methods.
A sequential design of minimax control and its limiting description are considered. The results of numerical experiments and Monte Carlo simulations are given.
Keywords: Two-armed bandit problem, Minimax and Bayesian approaches, Parallel processing, Partial differential equations

Real time Reliability of Fiscal stimulus – Comparison with USA, UK and Japan

Yasuyuki Komaki
Nihon University, College of Economics, Japan

There are many earlier theoretical and empirical studies for fiscal policy, as we have discussed the effect of fiscal stimulus since long ago. Nevertheless much estimation, we don't have the consensus for effect of fiscal stimulus in qualitative and quantity analysis because it is difficult to estimation. Additional above problems, we have crucial problems about

data revision. When we estimate the empirical evidence mainly using fiscal policy reaction functions, many macro-economic variables is revised after preliminary variables were published. The necessity for fiscal policy is decided by the preliminary variables at that time. Cimadomo (2008) shows the use of updated observations would point to a pro-cyclical fiscal policy stance in industrialized countries.
In this paper, we estimate the reliability of estimation on fiscal multiplier. After larger-scale fiscal stimulus, many earlier studies indicated that effect of fiscal stimulus has fallen. But whenever we estimate the fiscal multiplier, we have different results. Whichever we use, we find different results, as follow;
1) Government base: Multiplier did not go down after 90th clearly.
2) IG base and chain-linked data: Multiplier becomes smaller.
3) Continuous effect; In IG base, the effects of fiscal stimulus becomes stronger than first year. But in Government base the effects are smaller
On the other hand, we estimate the multiplier in USA and UK. The multiplier in USA changes after 1996/1Q estimation. In the comprehensive revision of January 1996, the BEA switched from reporting fixed-weight aggregation methods to chain-weight methods in its headline measures. We find same situation in Japan. Multiplier gets bigger after shift to chain-linked (2009/3Q). After introduction of chain-linked, investments including rapid change of commodity cycle like PC with IT have been fluctuation in price and volume. It may cause to the different results of fiscal multipliers.
But we find there is rarely difference before and after chain-weight methods in UK. Before shift to ESA 95, multiplier is significant larger. At least, the change of data property affects the empirical results of fiscal multiplier.
Key Words: Real Time, fiscal multiplier estimation, Reliability

Modelling relations between returns of financial investments using perturbed of copulas*

Jozef Komorník[1], Magda Komorníková[2], Jana Kalická[2]
[1]Faculty of Management, Comenius University, [2]Faculty of Civil Engineering, Slovak University of Technology, Slovakia

Copulas have obtained a growing interest in the last two decades, especially because of their numerous applications. Fitting of an appropriate copula to real data is one of major tasks in application of copulas. For this purpose, a large buffer of potential copulas is necessary, preferably parametric families of copulas. Once we know approximately a copula C appropriate to model the observed data, we look for a minor perturbation of C which fit better than C itself.

In our paper Mesiar, R., Komorník, J., Komorníková, M.: Modification of bivariate copulas, submitted to Fuzzy Sets and Systems, the next modification method applicable for any copula C was presented.

Theorem Let C: $[0, 1]^2 \to [0, 1]$ be a copula and define a perturbation $H_\lambda^C : [0,1]^2 \to [0,1], \lambda \in [0,1]$ by

$$H_\lambda^C(u, v) = \lambda \left(u - C(u, v)\right)\left(v - C(u, v)\right).$$

Then $C_{H_\lambda^C} : [0,1]^2 \to [0,1]$ given by

$$C_{H_\lambda^C}(u, v) = C(u, v) + H_\lambda^C(u, v) \tag{1}$$

is a copula for each $\lambda \in [0, 1]$ and any copula C.

We have investigated the relations between 4 selected countries' (USA, Australia, Japan and UK) daily returns of the REIT (Real Estate Investment Trust) indexes in different time periods, determined by the recent global financial markets crises (July 1, 2008 - April 30, 2009). We have applied the fitting by copulas to the residuals of ARMA-GARCH filters. We considered models from strict Archimedean copulas (Joe, Frank, Clayton and Gumbel) families and their mixtures with corresponding survival copulas (that have been applied e.g. in Patton's paper in International Economic Review 47, 2, 2006) as well as their modifications given by (1). For selecting the optimal models we have applied the Kolmogorov – Smirnov – Anderson - Darling (KSAD) test statistic (for which we also constructed a GoF simulation based test).

* Supported by grants APVV-0073-10 and VEGA 1/0143/11.

Adaptive access control in the server with a finite number of processors

Konovalov M.G.
The Institut of Informatics Problems of the Russian Academy of science(IPI RAN), Russia

The report proposes a solution to the following problem of congestion control. The computing complex (server) received a job from a random flow. The server contains a finite number of processors. On each processor can run at a time one job. The time of job execution is random. The server must decide every time whether it should take on this job or not. Each job, that was adopted at the service generates income R. However, the job has a deadline and if the task is not completed within the specified term, it is subject to a fine P > R. The server cannot accumulate too large queue because this can result to potentially large fine. So it has to reject part of the jobs and lose income.

We must find such a strategy of access, by which the average limit income will be maximal.
In the adaptive formulation of the problem it is assumed that the parameters of input flow of tasks, as well as the distribution of job times and deadline are unknown. Thus, the strategy must be based on the available observations, and to be able to self learn by the results of these observations.
Non-adaptive statement of the problem is considered in [1]. This problem is closely connected with the problem of overload control that occurs in the so-called SIP server networks [2].
Key Words: access control, adaptive strategy, job flow
References
1. Konovalov M. G. Ob odnoj zadache optimal'nogo upravlenija nagruzkoj servera [About one task of overload control] // Informatika i ee primenenija [Informatics and applications], 2013. V. 7. Iss. 4. Pp. 68–77 (in Russian).
2. Abaev P. O., Gaidamaka Y. V., Pechinkin A. V., Razumchik R. V., Shorgin S. Ya. Simulation of overload control in SIP server networks // ECMS 2012: Proceedings of the 26th European Conference on Modelling and Simulation. – Koblenz, Germany, 2012. P. 533–539.

Two-Stage Kalman Filtering for Discrete Systems Using Nonparametric Algorithms*

Gennady M. Koshkin, Valery I. Smagin
Tomsk State University, Russia

The paper addressed the filtering problem with using nonparametric algorithms for discrete stochastic systems by unknown input. The two-stage algorithm on the base of Kalman filtering and nonparametric estimator for systems with unknown input is designed and explored. Examples are given to illustrate the usefulness of the proposed results in comparison with the known methods.
Keywords: Kalman filter, Nonparametric estimator, Two-stage filtering algorithm, Unknown input

*Supported by Russian Foundation for Basic Research, projects no. 13-08-00744, and 13-08-98027

Monte-Carlo Reliability Evaluation of the Ring Detector based on Heavily Masked Normalized Correlation

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

A number of general-purpose ring and circle detectors are known. Typically, used techniques are template matching and Hough transform. However, ring detectors described in the literature were found impractical for the real-life implementation of the camera-based Instant Feedback System (IFS). Goal of the IFS is to collect answers of the students to the multiple-choice questions during the lecture. In the frames of the camera-based IFS, students answer to the specific multiple-choice question by presenting to the camera a specially designed IFS cards. In one preferred design IFS card contains bounding black ring, a plurality of internal markers and 2D barcode (unique for every student). In the human-friendly IFS card design, 2D barcode is a two-digit number specifying number of the student in the class list. Orientation of the IFS card specifies number of the selected answer. Human observer analyzing the image of the class easily recognizes bounding black ring and the number inside the ring for any possible orientation. For the computer the problem of recognition of the plurality of the IFS cards is not trivial because sizes of the bounding rings and digits are different. Additionally, most of the rings and digits in the periphery of the real-life image are directionally distorted, making standard template match approach too slow and non-reliable. To cope with this problem, standard Normalized Correlation template-matching algorithm was modified by adding the mask hiding the markers and digits inside the ring. At the first step, bounding rings are to be found, then, markers inside the ring are to be found and used to evaluate card orientation. It can be expected, that number of templates needed will be significantly smaller. In order to evaluate reliability of the proposed algorithm, special software Monte-Carlo simulator was created. The first part of the simulator creates image containing background image (for example, image of the real class) and a plurality of IFS cards in the pseudo-random positions and orientations. Additionally, pseudo-random directional distortions and uneven noisy illumination are applied. Boolean flag specify if cards overlap is allowed or forbidden. Second part of the Monte-Carlo simulator attempts to recognize IFS cards in the image in test. Monte-Carlo simulation results show that in case of non-overlapped cards recognition error is less than 1%, which can be considered as adequate for the real-life camera-based IFS. Developed approach can be used to speed-up recognition in the other practically interesting cases, for example, for the traffic signs recognition.

Keywords: Image Processing, Ring Detector, Normalized Correlation, IFS, Monte-Carlo simulation

Asymptotic Properties of Estimating Parameters of Intensity Function and Maintenance Effect

Makram Krit
Institut Supérieur d'administration des Entreprises, 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.

On the extended Orey index and its estimation

Kęstutis Kubilius
Vilnius University, Institute of Mathematics andInformatics, Lithuania

For real mean zero Gaussian process with stationary increments Orey [2] (see also [1]) suggested following definition of index.

Definition 1. Let X be a real-valued mean zero Gaussian stochastic process with stationary increments and continuous in quadratic mean. Let σ_X be the incremental variance of X given by $\sigma_X^2(h) = E[X(t+h) - X(t)]^2$ for $t,h \geq 0$. Define

$$\hat{\beta} := inf\left\{\beta > 0: \lim_{h\to 0} \frac{h^\beta}{\sigma_X(h)} = 0\right\} = \limsup_{h\to 0} \frac{\ln \sigma_X(h)}{\ln h} \qquad (1)$$

and

$$\hat{\beta}^* := sup\left\{\beta > 0: \lim_{h\to 0} \frac{h^\beta}{\sigma_X(h)} = +\infty\right\} = \liminf_{h\to 0} \frac{\ln \sigma_X(h)}{\ln h} \qquad (2)$$

If $\hat{\beta} = \hat{\beta}^*$ then X has the Orey index β_X.

If Gaussian process with stationary increments has Orey index then almost all sample paths satisfy a Hölder condition of order γ for each $\gamma \in (0, \beta_X)$. For fractional Brownian motion index $\beta_X = H$.

The purpose of this talk is to give an extension of the definition of the Orey index for the second order stochastic processes which may not

have stationary increments. Our goal is to estimate the Orey index for Gaussian process from discrete observations of its sample paths and consider the asymptotic properties of its estimator. Moreover, the exact confidence intervals for the Orey index are obtained.

References

1. Norvaiša, R., and Salopek, D.M., ``Estimating the Orey Index of a Gaussian Stochastic Process with Stationary Increments: An Application to Financial Data Set'', *Canadian Mathematical Society Conference Proceedings*, 26, 353-374 (2000).
2. Orey, S., ``Gaussian sample functions and the Hausdorff dimension of level crossings". *Z. Wahrsch. verw. Gebiete*, 15, 249-256 (1970).

Discrete observation of a continuous time semi Markov model for HIV control – Modelling the quality of life through rewards

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

The aim of the present paper is twofold. First, we review a continuous time non-homogeneous semi Markov model for HIV control and we apply an algorithm for data simulation analysis which we run to provide the data for transitions and the sojourn times of the corresponding visited states. After the simulation process three different models are developed and validated for the discrete observation of the simulated continuous process and an estimation method is applied to get the respective distributions. Secondly, a non homogeneous semi Markov reward model is introduced by reassigning the state space in order to include substates relative to the quality of life status of patients, for indexing through rewards the quality of life for patients of all health states. Means and higher order moments of rewards for periods of times for the patients are calculated as measures indicating quality of life status. Finally, the above results are illustrated numerically with synthesized data.

A multivariate SDE model for death rates: application to the Portuguese population data

Sandra Lagarto, Carlos A. Braumann, Dulce Gomes
Universidade de Évora, CIMA, Colégio Luís Verney, Portugal

We propose a new multivariate model to explain human mortality trends across time for all life span and for both males and females. In previous work [1], we assumed that there was an environmental randomness that affects in a similar way the crude death rates (CDR) associated to

individuals of different sexes and applied a bi-dimensional stochastic Gompertz model with correlated Wiener processes.
When observing the data, we noticed, however, that there are simultaneous mortality ups and downs, not only when comparing the time series by sex, but also when compared by age. This behavior suggests the application of a multidimensional model with a correlated noise structure.
Thus, assuming that each CDR, by age and sex, follows a one dimensional geometric brownian motion, we propose this new model, designated by multidimensional geometric brownian motion, in which the one-dimensional Wiener processes associated with different ages and sexes are correlated. This single model provides forecasts for CDR associated to individuals from age 0 (death occurred before the first complete year) to age 99 of both sexes, which is a huge advantage compared to the individual modeling of each series. The parameters are estimated by maximum likelihood, using numerical methods.
Keywords: Mortality rates, cross-sectional analysis, multivariate SDE models, correlated Wiener processes.
References
[1] Lagarto, S., Braumann, C.A. (2013) Modeling human population death rates: a bi-dimensional stochastic Gompertz model with correlated Wiener processes. Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications. Series Studies in Theoretical and Applied Statistics, XVII. Subseries Selected Papers of the Statistical Societies. Pacheco, A.; Oliveira, R.; Santos, R. (eds.), Springer, Berlin. (In print)

A Simulation Optimization Approach Based NSGA-II Algorithm for Solving a Queue Planning Problem

Nouha Lahiani[1,2], Yasmina Hani[1], Abdelfatteh Triki[2], Abderahman El-Mhamedi[1]
[1]LISMMA, Laboratoire d'Ingénierie des Systèmes Mécaniques et des MAteriaux, University of Paris VIII, France, [2]Laboratoire ARBRE- Institut supérieur de Gestion de Tunis,

In the current business environment, the manufacturing companies should be more reactive, flexible and competitive. These objectives can be realized by improving production and maintenance systems.
In this study, we consider a queuing problem of the maintenance interventions request (IR). In Fact, the bad management of the IR can impact greater costs.
In this paper, we propose a new perform approach for solving a maintenance process problem.

For this purpose, a combined simulation and non dominated sorting genetic algorithm II (NSGA-II) optimization model is developed.
Our simulation model is performed using FLEXSIM® software
The project does not deal with a classical simulation model which can optimize its results. We propose here an interactive decision tool. The qualifications of operators, their availability and the urgent degrees of interventions are considered. The NSGA-II and simulation models operate in parallel over time with interactions. That is in order to optimize the systems' performances; by choosing the best queues' scheduling policy. The proposed tool is based on multiobjective optimization problem described by three performances indicators: the number of the breakdown repaired, the stopping time of machines and the repairing time.
Applying this technique on an industrial case study, we show that our proposed approach is more effective in detecting real faults than existing alternatives.
Our approach can be extended to cover other domains and other types of simulation models
Keywords: Simulation based optimization, NSGA-II Algorithm, Case Study

Sequential estimation has two steps by the Bayesian approach

RIABI LAKHDAR
Oran, Algeria

There is a sequential estimation has two stages of averaging two people close by Bayesian introduce the quadratic Bayes risk
Key Words: séquentielle, risque de bayes, martingale, la famille exponentielle, temps d'arrét

Approach For Crew Pairing Problem With Resource Constraints

A. LAMAMRI[1], H. AIT HADDADENE[2]
[1]University of Khemis Miliana, [2]Laboratory LAID3, USTHB, Algeria

The crew pairing problem with resource constraints is solved using a mixed linear programming method. The problem is to cover cost of all flights of the company. Given the large size of the problems encountered in industry, these models are solved by an approach based on column generation that can handle implicitly all feasible solutions and a master problem determining the best solution. We present a new approach to improve the acceleration of the method of column generation problem for the construction crew rotations, it is projected in each arc, the resources

on a vector of dimension less using a Lagrangean relaxation algorithm to determine the coefficients of the projection arc combined with an algorithm for re-optimization, well generates a sub-set of complementary solutions to the master problem. Experimental results on real-world instances of the problem show that the proposed approach is capable of producing heuristic solutions of better quality than those obtained by these previous approaches, and of solving some small-size instances to proven optimality.
Key Words: optimization, column generation, crew pairing problem

Life Expectancy and Modal Age at Death in Selected European Countries in the Years 1950-2012

Jana Langhamrová, Kornélia Cséfalvaiová, Jitka Langhamrová
University of Economics, Czech Republic

At present, the majority of developed countries deal with the phenomenon of population ageing. This ongoing process is primarily caused by the increasing life expectancy at birth. Length of life is usually expressed by the indicator of life expectancy at age x. The values of life expectancy and modal age at death are different from the view of time evolution. This is particularly because life expectancy is the average age of deceased persons in the stationary population, whereas modal age at death is the most common age at death. The aim of this article is to analyse the trends in the development of the death rates in selected European countries using various methods for the death rates compensation. By means of data from the Human mortality database life expectancies and modal ages at death in selected countries will be calculated and compared. The purpose of this paper is to highlight the changes in the trend and dynamics of the life expectancy at birth and to compare its progress with the trend and dynamics of the modal age at death. By comparing the evolution of life expectancy at birth, life expectancy at age 65 and modal age at death, it is visible that modal age at death is not increasing as rapidly as life expectancy at age x. It is necessary to compensate the values of modal age at death or to use a more accurate calculation applying the Gompertz-Makeham function. Furthermore, there is a noticeable difference in the development of Western and Eastern Europe.
Keywords: Population ageing, Life expectancy at birth, Modal age at death

Stochastic Response Surface Methodology – a study on polynomial chaos expansion

Conceição Leal[1], Teresa Oliveira[1,2], Amílcar Oliveira[1,2]
[1]Department of Sciences and Technology, Open University, [2]CEAUL, Portugal

Risk analysis has assumed a crucial relevance over the past few years, particularly in dynamical systems with increasing complexity. Thanks to the recent technological advances, the use of simulation techniques became current to estimate models allowing predicting systems' behaviors for the probability of occurrence of a specific event and the consequences of this occurrence. Uncertainty associated with the simulation, either in model parameters or in experimental data, reveals its quantification as a prerequisite in probabilistic risk assessment.
The computational costs of numerical simulation are often very high, thus the use of metamodels arises as a pressing necessity. The Response Surface Methodology (RSM) is known to be a suitable tool, both for the estimation of metamodels on the behaviors of systems and risk assessment, as for the quantification of uncertainty.
Stochastic Response Surface Methodology was developed as a conceptual extension of the traditional RSM, to approximate model inputs and outputs in terms of random variables, such as standard normal variables, by a polynomial chaos expansion.
The objective of the methodology is to reduce the number of model simulations required for adequate estimation of uncertainty, as compared to conventional Monte Carlo methods.
In this work we review of this methodology and of some applications, and we present some examples using R-software.
Keywords: Uncertainty, Stochastic Response Surface Methodology, Risk Analysis, Polynomial Chaos.
Ackowledgments: This research was partially sponsored by national founds through the Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT under the projects (PEst-OE/MAT/UI0006/2011 and PEst-OE/MAT/UI0006/2014).

References

1. Balakrishnan, S., Roy, A., Ierapetritou, M. G., Flach, G. P., Georgopoulos, P. G.: Uncertainty reduction and characterization for complex environmental fate and transport models: An empirical Bayesian framework incorporating the stochastic response surface method. Water Resources Research, 39(12) (2003)
2. Bastug, E., Menafoglio, A., Okhulkova, T.: Polynomial Chaos Expansion for an Efficient Uncertainty and Sensitivity Analysis of Complex Numerical Models. Conference Paper. ESREL 2013, Amsterdam, Netherlands (2013)
3. Isukapalli, S. S., Roy, A., Georgopoulos, P. G.: Stochastic response surface methods (SRSMs) for uncertainty propagation: application to

environmental and biological systems. Risk analysis, 18(3), 351-363 (1998)
4. Isukapalli SS. An uncertainty analysis of transport transformation models. Ph.D. Thesis, New Brunswick, New Jersey: The State University of New Jersey; 1999.
5. Isukapalli SS, Roy A, Georgopoulos PG. Stochastic response surface methods for uncertainty propagation: application to environmental and biological systems. Risk Anal 1998;18(3):351–63.
6. Li, D., Chen, Y., Lu, W., & Zhou, C. (2011). Stochastic response surface method for reliability analysis of rock slopes involving correlated non-normal variables. Computers and Geotechnics, 38(1), 58-68.
7. Oladyshkin, S., Nowak, W.: Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliability Engineering and System Safety, 106, 179-190 (2012)
8. R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2012)
9. Xiu, D., Karniadakis, G.E.: Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. Computer Methods in Applied Mechanics and Engineering, 191(43), 4927-4948 (2002)
10. Xiu, D., Karniadakis, G. E.: Modeling uncertainty in flow simulations via generalized polynomial chaos. Journal of Computational Physics, 187(1), 137-167 (2003a)
11. Xiu, D., Karniadakis, G. E.: A new stochastic approach to transient heat conduction modeling with uncertainty. International Journal of Heat and Mass Transfer, 46(24), 4681-4693 (2003b)

Error bounds on practical approximation for two tandem queue with blocking and non-preemtive priority

Ouiza Lekadir, Djamil Aïssani
Research Unit LaMOS (Laboratories of Modelization and Optimization of Systems), Béjaia University, Algeria

Our work deals with the stability of the model [M2/G2/1 → •/G/1/1] tandem queue, with preemptive priority. This model is known to be analytically difficult to analyze due to the preemptive priority and the blocking features. However, this tandem queue is a good model for different fragments of queueing networks. It plays an important role for the validation of different decomposition algorithms designed for investigating more general queueing networks. The investigation of these queueing networks are usually solved only through approximations. It is therefore very important to justify these approximations and estimate the resulting error.

Thus, in this work, once the strong stability of our model established, we approximate the characteristics of the nominal model [M2/G2/1 → •/G/1/1] tandem queue, with preemptive priority, by those of the classical perturbed model [M/G/1 → •/G/1/1] tandem queue, when the arrival intensity of the priority stream is sufficiently small. This classical queueing network is simpler and more exploitable.
Moreover, explicit small error bounds will be obtained for the tail probabilities deviation between the nominal model and the perturbed model.
Key Words: Queueing networks, Tandem Queue, preemptive priority, Strong stability, Markov chain, Approximation, Inequalities of stability.

The incidence of cancer

C.T. Lenard[1], T.M. Mills[2], R.F.G. Williams[3]
[1]Mathematics and Statistics, La Trobe University, Australia, [2]Loddon Mallee Integrated Cancer Service, Australia, [3]School of Economics, La Trobe University, Australia

There is widespread interest in measuring the risk of being diagnosed with cancer. Internationally, especially in developed countries, governments collect and use data on the incidence of cancer for strategic planning to ensure that the nation has the resources that will be required to deal with the disease. Incidence data can also be used to assess the effectiveness of public health campaigns. However, there are several measures for quantifying the incidence of cancer. This paper examines the cumulative incidence rate. In this paper, we present a review of the method for estimating the cumulative incidence rate of cancer, and comparing these rates in two populations. We discuss the connection between the cumulative incidence rate and the cumulative risk of being diagnosed with cancer by a certain age, and provide details of the underlying mathematical ideas.

Are credit ratings time-homogeneous and Markov?

Pedro Lencastre[1,2], Frank Raischel[3], Pedro G. Lind[4], Tim Rogers[5]
[1]ISCTE-IUL, Portugal, [2]Mathematical Department, FCUL, University of Lisbon, Portugal, [3]Instituto Dom Luiz, University of Lisbon, Portugal, [4]ForWind and Institute of Physics, University of Oldenburg, Germany, [5]Centre for Networks and Collective Behaviour, Department of Mathematical Sciences, University of Bath, UK

We introduce a simple approach for testing the reliability of homogeneous generators and the Markov property of the stochastic

processes underlying empirical time series of credit ratings. We analyze open access data provided by Moody's and show that the validity of these assumptions - existence of a homogeneous generator and Markovianity - is not always guaranteed. Our analysis is based on a comparison between empirical transition matrices aggregated over fixed time windows and candidate transition matrices generated from measurements taken over shorter periods. Ratings are widely used in credit risk, and are a key element in risk assessment; our results provide a tool for quantifying confidence in predictions extrapolated from these time series.
Keywords: Generator matrices, Continuous Markov processes, Rating matrices, Credit Risk.

Bridging data and knowledge through a simple stochastic method

Pedro G. Lind
ForWind and Institute of Physics, University of Oldenburg, Germany

We present recent developments in a stochastic method that enables one to extract directly from sets of measurements the underlying governing equations. In particular, for a given data set we will derive the stochastic equation that statistically describes its evolution. We will also describe the physical inter-pretation of the coefficients defining the determinist and stochastic contributions separately. This method is parameter free and can be applied not only for the well-behaved Markov processes but also in cases where the data is spoiled with strong measurement noise. Further it can be implemented to single data series as well as to multivariate data, and it can be even adapted to non-stationary data sets. We will give concrete examples of how the method is applied to specific cases in geophysics, energy, neurosciences and economics.
Keywords: Stochastic Method, Data Analysis, Finance, Neurosciences, Wind Energy.

Competing Risks Modeled by Phase Type Distributions

Bo H. Lindqvist
Norwegian University of Science and Technology, Norway

Phase-type distributions represent the time to absorption for a finite state Markov chain in continuous time. The simplest examples are mixtures and convolutions of exponential distributions. The class of phase-type distributions is both flexible and conceptually simple to work with. Further, it is essentially no loss of generality to work with phase-type distributions, since any lifetime distribution can, at least in principle, be approximated arbitrarily close by a phase-type distribution. Phase-type distributions have received much attention in applied probability, in particular queuing theory. Here they generalize the celebrated Erlang distribution. In the present work we extend the phase-type methodology to the case of competing risks. The basic ingredient in a competing risks phase-type model is a finite state Markov chain in continuous time with more than one absorbing state, each absorbing state corresponding to a particular risk. Expressions for cause specific hazard functions, cumulative incidence functions etc. can now be given in terms of the transition matrix of the underlying Markov chain. Of particular interest is model estimation from (possibly censored) competing risks data. It will be indicated how estimation methods via the EM algorithm and via the MCMC method (respectively, Asmussen et al.,1996; Bladt et al., 2003) can be performed also in the case of more than one absorbing state. Successful applications of the MCMC method to phase type competing risks are reported in recent work by Laache (2014).
Key Words: Phase type distribution; Markov Chain; Competing risks; EM algorithm, Markov chain Monte Carlo

References

1. Asmussen, S., Nerman, O. and Olsson, M. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics, 23, 419-441 (1996).
2. Bladt, M., Gonzalez, A. and Lauritzen, S.L. The estimation of Phase-type related functionals using Markov chain Monte Carlo methods. Scandinavian Actuarial Journal 2003(4), 280-300.
3. Laache, C.H. Phase-type inference on competing risk models using Markov Chain Monte Carlo methods. Master thesis. Norwegian University of Science and Technology. 2014.

One-Way Repeated Measures ANOVA in the Study of the Lower Limbs' Explosive Power Level – a study with secondary education students

Domingos J. Lopes da Silva[1], Teresa Paula C.A. Oliveira[2], Amílcar Manuel R. Oliveira[3]

[1]Escola Superior de Educação de Fafe, Escola Secundária/3 de Barcelinhos, [2]Universidade Aberta, Centro de Estatística e Aplicações da Universidade de Lisboa, [3]Universidade Aberta, Portugal, Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

Objective: the present study intended to assess the explosive power of lower limbs (EPLI) of secondary school students.
Methods: the aleatory sample is constituted by 39 boys and 72 girls, of ages 15-16, students of the Secondary School of Barcelinhos (Barcelos, Portugal). The EPLI was assessed through the standing long jump test (jump performance, horizontal impulsion). The assessment focused on 4 moments of the school year: the beginning (M1), the end of 1st term (M2), the end of 2nd term (M3), and the end of the year (M4). In statistical terms, the significance level was of 5%, and all calculations were realized in SPSS 20. The statistical procedures used were the mean, standard deviation, minimum and maximum. The normality was verified by the Kolmogorov-Smirnov test, with Lilliefors correction. The variance homogeneity was assessed by Levene's test. The comparison between moments was made by one-way repeated measures ANOVA. The circularity of covariances matrix in endogenous variables was verified by Mauchly's sphericity test. In case of no sphericity violation, the Huynh-Feldt approximation was adopted. The post-hoc was performed by Bonferroni's test.
Results: the EPLI doesn't increase uniformly among the adolescents; while the boys improve, the girls stabilize.
Conclusion: the girls don't increase their performance throughout the school year, maintained practically unalterable EPLI values; the boys significantly improve the EPLI index, evidencing higher levels in the last moment (M4). Perhaps, the higher involvement of boys in activities that made use of strength is at the base of the results obtained.
Key-words: adolescents, standing long jump, explosive power of lower limbs, one-way repeated measures ANOVA

Ackowledgments: This research was partially sponsored by national founds through the Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT under the projects (PEst-OE/MAT/UI0006/2011 and PEst-OE/MAT/UI0006/2014).

A robust coefficient of determination for heritability estimation in genetic association studies

Vanda M. Lourenço[1], Paulo C. Rodrigues[2], Ana M. Pires[3]
[1]FCT/UNL and CMA-FCT/UNL, [2]CMA-FCT/UNL, [3]CMA IST/UTL and CEMAT-IST/UTL, Portugal

Heritability is key in plant studies to help achieve better yield and other agronomic traits of interest. In candidate gene studies regression models are used to test for associations between phenotype and candidate SNPs. SNP imputation guarantees that marker information is complete and the data are balanced. So both the coefficient of determination, R^2, and broad-sense heritability are equivalent. However, when the normality assumption is violated, the classical R^2 may be seriously affected. Recently two R^2 alternatives with good properties were proposed for the linear mixed model. We evaluate their performance under contamination and step forward a robust version of these coefficients assessing their adequacy for heritability estimation via simulation. An example of application is also presented.

Keyworks: Robust coefficient of determination; Single nucleotide polymorphism; Association studies; Robust regression.

Acknowledgements: This work received partial financial support from Portuguese National Funds by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through projects PTDC/MAT-STA/0568/2012 and PEst-OE/MAT/UI0297/2014 (Centro de Matemática e Aplicações).

Asian Options, Jump-Diffusion Processes on a Lattice, and Vandermonde Matrices

Karl Lundengård, Carolyne Ogutu, Sergei Silvestrov
Mälardalen University, Sweden

Asian options are options whose value depends on the average asset price during its lifetime. They are useful because they are less subject to price manipulations. We consider Asian option pricing on a lattice where the underlying asset follows Merton–Bates jump-diffusion model We describe the construction of the lattice using the moment matching technique which results in an equation system described by a Vandermonde matrix. Using some properties of Vandermonde matrices we calculate the jump probabilities of the resulting system. Some conditions on the possible jump sizes in the lattice are also given.

Key Words: Jump-diffusion process, lattices, Vandermonde matrix, Asian options, option pricing

Statistical Design of Adaptive Control Charts for Linear Profile Monitoring

Maysa S. De Magalhães[1], Viviany L. Fernandes[2], Francisco D. Moura Neto[3]
[1]National School of Statistical Sciences, Brazilian Institute of Geography and Statistics (ENCE/IBGE), [2,3]Polytechnic Institute, State University of Rio de Janeiro (IPRJ/UERJ), Brazil

In some production processes the quality characteristics can be represented by profiles or linear functions. We propose adaptive control charts to monitor the coefficient vector of a simple linear regression model, once fixed parameter control charts are slow in detecting small to moderate shifts in the process parameters, that is, the intercept and the slope. A study on the performance of the proposed control charts was done, considering the average time until a signal.
Keywords: Linear profile, Adaptive control chart, Markov chain

Investigating of extreme value threshold estimation approaches

Justyna Majewska
University of Economics in Katowice, Poland

The determination of the threshold where the tails of a distribution begin remains a fundamental statistical problem. From a statistical point of view the threshold is not precisely defined. The paper reviews recent advances and some traditional threshold estimation approaches for extreme value applications. We investigate the consequences of using standard methods with mis-specified threshold and we discuss the problems that may occur when we do not take into account threshold uncertainty.
To gain insight into the behaviour of the different threshold estimation procedures simulated data from a known distribution are used. The major benefits and drawbacks of this parametric, nonparametric and semi-parametric approach are provided.

Set-valued and fuzzy stochastic differential equations

Marek T. Malinowski
Faculty of Mathematics, Computer Science and Econometrics
University of Zielona Góra, Poland

A typical feature of many real-world phenomena is uncertainty. This term is mostly understood as a stochastic uncertainty and methods of probability theory are utilized in its analysis. However, uncertainty can result from the second source: vagueness (sometimes called ambiguity, fuzziness, softness). It appears, for instance, when the data of considered systems are imprecise due to lack of precision of measuring instruments. This type of uncertainty is not stochastic. With the notions of set-valued stochastic differential equations and fuzzy stochastic differential equations we offer the new mathematical tools that can be appropriate in modeling dynamical systems subjected to two combined types of uncertainties, i.e. randomness and vagueness (multivaluedness, fuzziness), simultaneously. To this aim we will consider the set-valued and fuzzy stochastic integrals. We will present the existence and uniqueness theorems for solutions to the set-valued and fuzzy stochastic differential equations. We will consider the solutions with increasing and decreasing diameter of their values. Some examples of the set-valued and fuzzy stochastic differential equations will be solved.

References

1. Malinowski, M. T., On random fuzzy differential equations, Fuzzy Sets Syst. 160, 3152-3165 (2009)
2. Malinowski, M. T., Existence theorems for solutions to random fuzzy differential equations, Nonlinear Anal. TMA 73, 1515-1532 (2010)
3. Malinowski, M. T., Random fuzzy differential equations under generalized Lipschitz condition, Nonlinear Anal. Real World Appl. 13, 860-881 (2012)
4. Malinowski, M. T., Strong solutions to stochastic fuzzy differential equations of Itô type, Math. Comput. Modelling 55, 918-928 (2012)
5. Malinowski, M. T., Itô type stochastic fuzzy differential equations with delay, Systems Control Lett. 61, 692-701 (2012)
6. Malinowski, M. T., Some properties of strong solutions to stochastic fuzzy differential equations, Inform. Sciences 252, 62-80 (2013)
7. Malinowski, M. T., Approximation schemes for fuzzy stochastic integral equations, Appl. Math. Comput. 219, 11278-11290 (2013)
8. Malinowski, M. T., On a new set-valued stochastic integral with respect to semimartingales and its applications, J. Math. Anal. Appl. 408, 669-680 (2013)
9. Malinowski, M. T., Modeling with stochastic fuzzy differential equations, In: Mathematics of Uncertainty Modeling in the Analysis of Engineering and Science Problems (S. Chakraverty, Ed.), pp. 150-172, IGI Global, Hershey Pennsylvania (2014)

Steady states for non-Markovian stochastic synchronization

Anatoly Manita

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

Stochastic particle systems with synchronization-like interaction are a convenient mathematical tool for modelling synchronization phenomena in different application areas, see, e.g., [1] and [2]. We study a long time behavior of synchronization models with a large number of components under rather general assumptions about epochs of synchronizations between the components. Our goal is to obtain large scale approximations for probability distributions of distances between particles.

Keywords: Stochastic synchronization, multi-component systems, limit distributions, renewal process.

Acknowledgement.
The work is supported by Russian Foundation for Basic Research (Grant No. 12-01-00897).

References

[1] A. Manita, On Markovian and non-Markovian models of stochastic synchronization. Proceedings of The 14th Conference ASMDA, 2011, Rome, Italy, p. 886-893.

[2] A. Manita, Intrinsic space scales for multidimensional stochastic synchronization models. In "New Perspectives on Stochastic Modeling and Data Analysis", International Society for the Advancement of Science and Technology Publisher, 2014, Athens, Greece

Singular extremals in control problems for wireless sensor networks

Larisa Manita

National Research University Higher School of Economics, Institute of Electronics and Mathematics, Russia

Energy-saving optimization is very important for various engineering problems related to modern distributed systems, see, e.g., the survey [1] and references therein. We consider here a control problem for a wireless sensor network with a single time server node and a large number of client nodes, see [2]. The problem is to minimize a functional which accumulates clock desynchronization errors in the clients nodes and the energy consumption of the server over some time interval $[0,T]$. The control function $u=u(t)$, $0 \leq u(t) \leq u1$, corresponds to the power of the server node transmitting synchronization signals to the clients. For all possible parameter values we find the structure of extremal trajectories. We show that for sufficiently large u_1 the extremals contain singular arcs.
Keywords: Pontryagin maximum principle, bilinear control system, switching points, singular extremals

Acknowledgement.
The work is supported by Russian Foundation for Basic Research (Grant No. 13-08-00665).

References

[1]Sundararaman, B., Buy, U., Kshemkalyani, A.D.: Clock synchronization for wireless sensor networks: a survey. Ad Hoc Networks, 3, 3, 281–323, 2005

[2] Manita A., Clock synchronization in symmetric stochastic networks, Queueing Systems, 76, 2, 149-180, 2014

From deterministic to stochastic via characteristic systems

Ionela Marinela Marinescu
Department of Applied Mathematics, Bucharest University of Economic Studies, Romania

In this paper are presented some solutions for linear and nonlinear SPDEs obtained from deterministic hyperbolic equations involving stochastic perturbations generated by a finite set of vector fields which commute using Lie bracket multiplied by a vectorial Wiener process. In the linear case the hyperbolic equation comes from scalar higher order deterministic equations determined by a derivation. In the nonlinear case we start with a nonlinear deterministic equation and stochastic perturbation is generated by a finite set of nonlinear vector fields which commute using Lie bracket.
Acknowledgements. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, Project No. IDEI 303, code PN-II-ID-PCE-2011-3-0593.
Key Words: Stochastic characteristic systems, irregular perturbations of nonlinear (H-J) equations

Performance of combined models on multi-class discrete classification problems

Anabela Marques[1], Ana Sousa Ferreira[2], Margarida G. M. S. Cardoso[3]
[1]Barreiro College of Tecnology, Setúbal Polytechnic ,IPS, Portugal, [2]Faculty of Psychology, University of Lisbon, Portugal and UNIDE, [3]Dep. of Quantitative Methods of ISCTE - Lisbon University Institute, Portugal and UNIDE.

In this work we evaluate the performance of a combined model for multi-class Discrete Discriminant Analysis. We specifically consider a convex combination of the First-order Independence Model (FOIM) and of the Dependence Trees Model (DTM).
We use simulated data sets with different levels of the classes' separation and with different sample sizes (moderate and small sized samples are considered). In order to evaluate classification performance we resort to the percentage of correctly classified observations and also to the Huberty index – the calculation of both measures is based on cross-validation. The results obtained show the advantage of the proposed combined model when compared to FOIM and DTM and allow us to understand the influence of classes' separation and sample sizes on classification performance.
Keywords: Classification performance, Combining models, Dependence Trees model, Discrete Discriminant Analysis, First-order Independence model, Sample sizes, Separability.

References
1. Brito, I., Celeux, G. and Sousa Ferreira, A., Combining methods in supervised classification: A comparative study on discrete and continuous problems. REVSTAT - Statistical Journal, 4(3), 201-225 (2006).
2. Marques A., Sousa Ferreira, A. and Cardoso, M. Selection of variables in Discrete Discriminant Analysis. Biometrical Letters, 50(1), 1-14 (2013).

Modelling patient length of stay in using the discrete conditional phase-type model with decision trees

Adele H Marshall[1], Mariangela Zenga[2]
[1]Centre for Statistical Science and Operational Research (CenSSOR) Queen's University Belfast, UK, [2]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy

Discrete Conditional Phase-type models (DC-Ph) are a family of models capable of representing a skewed survival distribution as a Process component preceded by a set of related discrete variables that may be referred to as the Conditional component. The process component in the DC-Ph model comprises of a Coxian phase-type distribution which represents a (typically skewed) service or survival time distribution, equating to the time to absorption in a continuous time finite Markov chain. The conditional component, which precedes the process component, will in this paper, comprise of a decision tree that uses discrete variables about the patient to identify cohorts of patients based who survival behaves in a similar way. The approach is applied to patient length of stay data for elderly patients in hospitals in Italy. The DC-Ph model is fitted with a classification tree, the outcomes of which are fed into multiple Coxian phase-type distributions representing patient length of stay. The best fit DC-Ph model indicated the most relevant patient information when predicting patient outcome.

Time to degree and gender differences: the case of an Italian and Greek university

Adele H. Marshall[1], Mariangela Zenga[2], Aglaia Kalamatianou[3]
[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy, [2]Centre for Statistical Science and Operational Research (CenSSOR) Queen's University Belfast, UK, [3]Department of Sociology, Panteion University of Social and Political Sciences, Greece

By most objective standards, women in social sciences disciplines are doing better regarding time to degree. This paper, using data from a

number of Italian and Greek social sciences departments, documents the extent of this gender differential controlling for variables related to students' personal and educational characteristics. The analysis is based on a statistical methodology recently introduced by the authors of this work, that combines Coxian phase-type distributions decision trees and the Gini index, and has been proved useful for handling issues relating to time to an event, building also on the censored nature of the available data.

National Survey for Non-Communicable Disease Risk Factors and Injuries for Timor-Leste

João Soares Martins
Faculdade de Medicina e Ciências da Saúde, Universidade Nacional Timor Lorosa'e, Timor-Leste

Non-communicable diseases (NCDs) are group of conditions that covers cardiovascular diseases, cancer, mental health problems, diabetes mellitus, chronic respiratory disease and musculoskeletal conditions. A large proportion of NCDs are preventable which can be done through modifying behavioural risk factors such as tobacco use, unhealthy diet, lack of physical activity, and the harmful use of alcohol. These behavioural factors lead to four major metabolic risk factors: overweight/obesity, high blood pressure, raised blood sugar and raised blood lipids which in turn are responsible for four group of major NCDs – cardiovascular diseases, diabetes, chronic respiratory diseases and cancers. NCDs kill more than 36 million people per year and it also contributes to sizeable economic impacts on households, industries and societies through losses in income, productivity and capital formation.

Timor-Leste is facing a double burden of disease. Communicable diseases such as TB, Malaria and Dengue continue to pose a public health challenge. On the other hand, NCDs such as cardiovascular, chronic obstructive pulmonary diseases have emerged among the top ten causes of mortality. Hospital data in 2012 indicate that about 22 percent all deaths were due to NCD groups.

Information on NCD and Injury risk factors is essential to plan an evidence based National strategy for their prevention and control. Timor-Leste is in the process of formulating its national strategy for NCD and Injury prevention and control and fixing its own targets.

As the stated in the grand strategic development plan 2011-2030, the goal of health sector development by 2030 was that Timor-Leste would have a healthier population as a result of comprehensive, high quality health services accessible to all Timorese people. To achieve a healthier population in this country, the role of NCDs and Injuries cannot be ignored; they play a central part in this.

The purpose of the proposed national survey is to establish a baseline for Timor-Leste on NCD and injury situations for understanding the magnitude of the NCD and Injury problems which will contribute to effective monitoring and evaluation of NCD trends. This will also help to develop a comprehensive national program on NCD and injury prevention and control in Timor-Leste. The survey will follow the WHO's eSTEPS survey methodology and will be expected to contribute to the WHO's global monitoring of NCD status.

On a family of Bivariate Compound Generalized Power Series Distributions

Leda D. Minkova
Sofia University, Bulgaria

The Inflated--parameter Power Series distributions (IPSD) was introduced by Minkova (2002), as a compound Power Series distributions (PSD) with geometric compounding distribution. The lack of memory property of the geometric distribution leads to some useful properties of the defined family of istributions.
In these notes I introduce a family of compound GPSDs with bivariate geometric compounding distribution. The probability mass function, recursion formulas and some properties are given. The particular cases of bivariate compound binomial, negative binomial and logarithmic series distributions are analyzed in detail.

A Coxian phase-type model with a hidden node for patient quality of care

Hannah Mitchell[1], Adele H Marshall[1], Mariangela Zenga[2]
[1]Centre for Statistical Science and Operational Research (CenSSOR) Queen's University Belfast, UK, [2]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy

Healthcare systems, the running of them and the quality of care delivered by them, have come under increasing attention in recent years. Healthcare managers have come under pressure to make sure that the hospitals are delivering the best quality of care available but at the same time effectively managing a budget. Quality of care can be defined in many ways, making its measurement and incorporation into scientific study difficult. One way to attempt to model the concept of quality of care is to use a hidden markov model. Modelling patient flow in healthcare systems is considered vital in understanding the system's activity. This paper presents a Coxian phase-type model with a hidden node which will

attempt to model patient flow in healthcare incorporating into it the quality of care delivered by the hospital.

Multi-commodity Network Flow Problem with Arc failures

Hiroaki Mohri[1], Jun'ichi Takeshita[2]
[1]Shinjuku-Ku, Faculty of Commerce, Waseda University, [2]Research Institute of Science for Safety and Sustainability (RISS), National Institute of Advanced Industrial Science and Technology (AIST), Japan

In this presentation, we would like to address a recovery problem on a multi-commodity flow network $G(N,E,K)$ with arc failures. In natural disasters like big earthquakes e.g. the great East Japan Earthquake, we have to recover some infrastructure networks as road, railway, electric, telecommunication networks and so on. We suppose all information of the network during the operation and can get an optimal solution on the network problem without failures. Regarding arc failures, we suppose all arcs have the same distribution.
We would discuss on strategies of recovering network. On some special graphs which have good structure, we propose an algorithm to get a feasible solution in polynomial time. Also, we would like to show an improving algorithm for the solution. Basically, we decompose the original graph into $G(N_i,E_i,K_i)$. Then, we try to use solutions of decomposed graphs.
Keywords: Stochastic modeling, Graph Theory, Optimization, Reliability Theory
References
1. Colbourn, C.J., The Combinatorics of Network Reliablity. Cambridge University Press (1987)
2. Diestel, R., Graph Theory 2nd Edition. Springer-Verlag (2000).
3. Kennington, J.L, A Survey of Linear Cost Multicommodity Network Flows, Operations Research, Vol.26, No2 (1978)
4. Shier, D. H., Network Reliability and Algebraic Structure. Oxford University Press (1991)

Global limit theorems and their applications

Stanislav Molchanov
UNC Charlotte, USA

Let $X_1,...,X_n$ be i.i.d. random vectors with symmetric density $p(x0 = p(-x), x \in R^d$, finitely many moments (of the order at least 2) and power tails. Under some regularity conditions on the behavior of

$p(x), |x| \to \infty$ we prove the global limit theorem for the density $p_n(x)$ of the normalized sum $Z_n = (X_1 + ... + X_n)/\sqrt{n}$ in the spirit of the Linnik's results (see Ch. XIV of the well-known book by I. Ibragimov and Yu. Linnik).
The applications of the global limit theorems include the theory of the phase transitions for the homo-polymers with slowly decreasing potential and the phenomenon of the intermittency for the contact processes and population dynamics.

Stochastic Modeling in Sexually Reproducing Biological Populations Through Branching Models

Manuel Molina, Manuel Mota, Alfonso Ramos
Department of Mathematics, University of Extremadura, Spain

The aim of this contribution is to investigate appropriate mathematical models to describe the demographic dynamics in sexually reproducing biological population. Significant efforts have been made to develop models based on the assumption that both phases, mating and reproduction, are influenced by the number of couples in the population, see e.g. Ma (2006), Molina (2008, 2012) or Xing and Wang (2005). However, in many biological species with sexual reproduction, such phases rather than being influenced by the number of couples, they are affected by the number of females and males in the population. Mathematical models to describe the demographic dynamics of such population have not been studied in the literature about branching processes. The motivation behind this contribution is to help to solve this issue. To this end, we introduce a new class of two-sex branching models where several mating conducts between females and males and various reproductive behaviors are taken into account. Both phases, mating and reproduction, are influenced by the number of females and males in the population. Under a general parametric setting, we study some inferential questions about the main parameters affecting the reproduction phase. By considering the observation of the number of females and males up to a certain pre-set generation is reached, we determine Bayes estimators for such parameters. Also, with the purpose to determine the corresponding highest posterior density credibility sets, we propose a computational algorithm. As illustration, we present an application to salmon populations.
Keywords: Stochastic modeling, Branching models, Two-sex models, Bayesian inference, Population dynamics
References

1. Ma, S. “Bisexual Galton-Watson branching processes in random environments“, Acta Math. Appl. Sinica, 22, 419-428 (2006).
2. Molina, M., Jacob, C., and Ramos, A. “Bisexual branching processes with offspring and mating depending on the number of couples in the population”, Test, 17, 265-281 (2008).
3. Molina, M., Mota, M., and Ramos, A. “Two-sex branching models with random control on the number of progenitor couples, Methodol. Comput. Appl. Probab., 14, 35-48 (2012)
4. Xing, Y., and Wang, Y. “On extinction of a class of population-size dependent bisexual branching processes”, J. Appl. Probab., 42, 175-184 (2005).
*This research has been supported by the Gobierno de Extremadura (grant GR10118), the Ministerio de Economía y Competitividad of Spain (grant MTM2012-31235) and the FEDER.

Two-sex branching processes with mating and reproduction phases influenced by the number of females and males in the population

Manuel Molina, Manuel Mota, Alfonso Ramos
Department of Mathematics, University of Extremadura, Spain

This research is motivated by the problem of developing appropriate mathematical models to describe the probabilistic evolution of biological populations with sexual reproduction. To this end, several classes of branching processes have been developed, see for details the surveys by Hull (2003) and by Molina (2010). In such classes of branching processes, mating and/or reproduction are affected by the number of couples (female-male) in the population. However, the mating and reproduction phases are usually affected by the number of females and males in the population. This practical situation has not been considered in the literature on branching processes. In this paper, we introduce a new class of two-sex branching processes where, in each generation, both the mating and the reproduction phases are influenced by the current number of females and males in the population. For such a class of stochastic models, we investigate some probabilistic questions. We especially investigate the possible extinction of the population. First, we study conditions which guarantee the classical extinction-explosion property in branching process theory. Then, we investigate conditions for the almost sure extinction of the population or for its survival with a positive probability.
Keywords: Branching processes, Two-sex processes, Extinction probability, Population dynamics.

References
1. Hull, D.M. “A survey of the literature associated with the bisexual Galton-Watson branching process“, Extracta Math., 18, 321-343 (2003).
2. Molina, M. “Two-sex branching process literature”, Lectures Notes in Statistics, 197, 279-293 (2010).

Sommerfeld’s Integrals and Hallén’s Integral Equation in Data Analysis for Horizontal Dipole Antenna above Real Ground

Farid Monsefi[1], Milica Rančić[1,2], Sergei Silvestrov[1], Slavoljub Aleksić[2]
[1]Division of Applied Mathematics, UKK, Mälardalen University, Sweden, [2]Dept. of Theoretical Electrical Engineering, ELFAK, University of Niš, Serbia

Increase of the radiation power in different frequency bands during the last decades, has called for a study of harmful effects on the living organisms and electronic equipment of the radio frequency energy. An accurate determination of the near field strength, electric as well as magnetic, in the vicinity of higher-power transmitting antennas is necessary for assessing any possible radiation hazard. In that sense, it is of great importance to account for the influence of the finite ground conductivity on the electromagnetic field structure in the surroundings of these emitters. The estimation of this influence has been intensively studied, and a number of approaches has been applied in that sense, ranging from the exact full-wave based ones to different forms of approximate, less time-consuming, ones. Although the approximate methods introduce a certain level of calculation error, their simplicity is of interest in the electomagnetic compatibility (EMC) studies. For that reason, finding an approximate, but satisfyingly accurate method, applicable to wide range of parameters is often a goal of researches done in this field.
In this paper, the authors perform an analysis of a thin horizontal dipole antenna (HDA) above real ground of known electrical parameters. The approach is based on the electric-field integral equation method, and formulation of the Hallén’s integral equation (HIE). This equation is then solved for the current, which is assumed in a polynomial form, using the point-matching method (PMM). This way obtained system of linear equations involves improper Sommerfeld’s integrals, which express the influence of the real ground and are here solved approximately using simple, so-called OIA and TIA, approximations (one- and two-image approximations). Both types of approximations are in an exponential form, and therefore are similar to those obtained applying the method of images. It should be kept in mind that the goal of this approach is to develop approximations that have a simple form, whose application yields satisfyingly accurate calculations of the Sommerfeld`s type of

integrals, and are widely applicable, i.e. their employment is not restricted by the values of electrical parameters of the ground, or the geometry.
Thorough analysis is performed in order to observe the influence of different parameters of the geometry, and the ground, on current distribution and the input impedance/admittance of the HDA in a wide frequency range. Furthermore, the verification of the method is done by comparison to the exact model based on the full-wave theory, and experimental data. Obtained results indicate a possibility of applying the described methodology to inverse problem involving evaluation of electrical parameters of the ground (or detection of ground type change) based on measured input impedance/admittance of the antenna.
Keywords: Horizontal dipole antenna, Hallén's integral equation, Point-matching method, Polynomial current approximation, Real ground, Sommerfeld's integrals

The association of ethnicity and maternal education with mortality risk at young ages using indirect estimates. Brazil, 2000 and 2010

Mario F. G. Monteiro, Alba M. Zaluar
NUPEVI, State University of Rio de Janeiro, UERJ, Brazil

Main topic: Human population and mortality data and databases

Probability of premature death between the ages of 15 and 30 (15q15) is particularly important in studies on violence because external causes account for 72% of death in this age group.
The objective of this paper is to study the ethnic and mother's education risk differences of mortality among young people, estimating the probability of an adolescent that is 15 years old to die before reaching 30 years old (15q15). This information can be useful for implementing public policies that aim a reduction of mortality risk at young ages. The estimates have been done for five selected States of Brazil: Rio de Janeiro, São Paulo and Minas Gerais, in the Southeast Region, and Pernambuco and Bahia in the poorer Northeast Region.
Using the MortPak, program, developed by the UN Population Bureau, with information from the Demographic Censuses of 2000 and 2010 on children born alive and children surviving according to the mother's age (Method of Brass), were estimated the probability of death between the ages of 15 and 30 years old (15q15).
The results showed higher risks for the Black population than for the White, mainly in Pernambuco in 2000 (15q15 = 48/1000, meaning that 4,8% of young Black, aged 15 years old, will die before completing 30 years) and Bahia (15q15 = 34/1000, in 2000). Nevertheless, significant reductions of 15q15 from 2000 to 2010 were observed for Black and

White in five States. For the Black population in Pernambuco, the estimate of 15q15 was reduced from 48/1000 in 2000 to 26/1000 in 2010.
The association of maternal education with 15q15 will also be analyzed.
Keywords: Mortality at young ages, indirect estimates of mortality.

Temperature fluctuation analysis at ESPOL campus in Guayaquil after seven years

Daniel Moran-Zuloaga[1], Guillermo Soriano[2], Ruben Hidalgo[2]
[1]Escuela Superior Politécnica del Litoral ESPOL, [2]Laboratorio de Fuentes Renovables de Energía LABFRE, Ecuador

Guayaquil is the most populated city in Ecuador (~2 million hab.) and its citizens believe that temperature has increased lately due the Anthropogenic events. In order to tested, Escuela Superior Politécnica del Litoral University, ESPOL has been monitoring since 2008 the weather conditions of its campus and provided a time series data since 2008 until 2014 in which for this research we will mainly focused on the temperature. Therefore, a daily lecture of every 10 minutes was tested in order to evaluating from a linear model, Auto Correlation Function ACF, periodogram the behavior of the temperature in ESPOL campus. We found that the temperature pattern had a linear behavior with an average temperature of ±25° C. On the other hand, with the ACF we found that major fluctuations occurred in small intervals (<50000 lectures), mainly produced white noise. A periodogram with a power spectrum with a Fourier correlation provided it us a negative trend. In summary, with the time series analysis we can inferred that Guayaquil is experience higher fluctuations from temperatures lower than 20°C to higher than 35°C. Thus, this research disapproved that the temperature is increasing in the city. Instead, we think the temperature is stable with extreme events as white noise.
Key Words: Linear trend, Fourier correlation, Auto correlation function, Temperature, Guayaquil

Shortcoming in some Index: A Mathematical Modeling

Barun Kumar Mukhopadhyay
Retired Scientist, Population Studies Unit, Indian Statistical Institute Kolkata, India

Historically indices have perpetually been constructed to measure the digit preference error in the distribution of single year age by researchers, actuaries. Simultaneously process of modification has been

going on in parallel from time to time by them. All these have taken place because of methodological drawbacks or many constraints came on the way of constructing the same. As a result the process was continued in a fashion from long past, in the recent past and at present time too. Every researcher while doing any socio demographic project or even in the censuses of many countries, especially in the developing world at an initial stage use to focus on the data quality. As per the present context of the paper, obviously age reporting error is significant to manifest either in the census or project report based on the data either from census or sample. Usually Myers' index (1940), to a large extent, has been used in those records and reports. Many attempts have been made (King, (1915); Myers(1940); Bachi (1951); Ramachandran, (1965) and others) to modify the methodologies. The current paper is based on a mathematical model to show how the many defects at a time are eradicated in a single attempt. The methodology was applied on the single year age data of Indian census, 2001. The result arrived at has been compared with the actual calculation of the index and the appropriateness of the present technique has been discussed.
Keywords: Myers' index, Whipple's index, Digits, Digit Accuracies

Application of Stable Distributions to the analysis of GPS-derived telemetry data

Mutwiri R. M, Mwambi H, Slotow R, Vanak A. T
University of Kwazulu Natal, South Africa

Probabilistic approaches used in movement ecology applications attempt to capture the properties exhibited by the random variables describing the movement behaviour of the animal in heterogeneous environment. In a standard parametric approach, the underlying data is modeled using assumed probability distributions.
The model parameters are fitted to the observed data using empirical or maximum likelihood methods. The model is then used to make decisions on out-of-sample data for any potentially heavy tailed and skewed events. In practical applications, large data sets obtained from GIS telemetry studies of animal tracking exhibits heavier tails than Gaussian distributions and have non-zero skewness. The tail behaviour of such distributions offer valuable information for decision support and conservation management tools used by wildlife ecologist.
This paper introduces the application of stable distributions to a data set of step lengths acquired from an individual elephant tracked in Kruger National park in south Africa. Stable distributions are a class of probability distributions that generalize Gaussian distributions and which can accommodate heavy tails and skewness. The results show that stable distributions accurately describe tail events of real life data in an

accurate manner and can be used as an alerting tool in wildlife management.
Key Words: Heavy-tailed distribution; Stable distribution; Parameter estimation; GPS-telemetry; animal movement.

Genetic Data Base

Kakha NADIRADZE, Nana PHIROSMANASHVILI
BioNanoTechnology Research Institute, Georgia

Genetic algorithms have made a major contribution to optimization, adaptation, and learning in a wide variety of unexpected fields. Over the years, many excellent books in genetic algorithm optimization have been published; however, they focus mainly on single-objective discrete or other hard optimization problems under certainty. There appears to be no book that is designed to present genetic algorithms for solving not only single-objective but also fuzzy and multiobjective optimization problems in a unified way. Genetic Algorithms And Fuzzy Multiobjective Optimization introduces the latest advances in the field of genetic algorithm optimization for 0-1 programming, integer programming, nonconvex programming, and job-shop scheduling problems under multiobjectiveness and fuzziness. In addition, the book treats a wide range of actual real world applications. The theoretical material and applications place special stress on interactive decision-making aspects of fuzzy multiobjective optimization for human-centered systems in most realistic situations when dealing with fuzziness.
Key Words: Multiobjective Optimization, Genetic, Fuzzy

Forecasting of jack mackerel landings (Trachurus murphyi) in central-southern Chile through neural networks

LAURA NARANJO, ELEUTERIO YÁÑEZ, MARÍA ÁNGELA BARBIERI, FRANCISCO PLAZA
Pontificia Universidad Católica de Valparaíso, Chile

The performance of neuronal networks models in monthly landing forecasting of jack mackerel (Trachurus murphyi) in central-southern Chile (32ºS-42ºS) is assessed. Thus, monthly estimations for ten environmental variables, fishing effort and jack mackerel landings for the period 1973-2008 are used. A preliminary analysis is done in order to remove strongly correlated variables. Sea surface temperature and fishing effort are established as input variables, then, a non-linear cross correlation analysis is performed to estimate the lag between the input variables and jack mackerel landings. Two models are adjusted: model

one includes both training and testing cases are randomly selected using all data involved in the analysed period; for model 2, the data is divided into two time series: the first from 1973 to 2002 used for training, and the second between 2003 and 2008 used for validation. The external validatio! n process for model 1 showed an explained variance of 92%, with a standard forecasting error of 30%. The explained variance for model 2 was 81%, with a standard forecasting error of 38%. Finally, the sensitivity analysis for both models showed the fishing effort as the most influential variable to jack mackerel landings, which presents functionality depending on anthropogenic effects rather than environmental conditions.

Key Words: Landings, Jack Mackerel, Forecasting, Fishing Effort, Environment, Neural Networks.

The Equivariance Criterion in Statistical Prediction

Tapan Nayak, Haojin Zhou
Department of Statistics, George Washington University, USA

This paper offers a general formulation of the basic logic of equivariance in the context of a point prediction problem. We propose a framework that allows the set of possible predictions and also the loss function to depend on the data and then explore the structure and properties of transformation groups that are relevant for applying the functional and formal equivariance principles. We define loss invariance and predictive equivariance appropriately and discuss their ramifications. We describe a structure of equivariant predictors in terms of maximal invariants and present a method for deriving minimum risk equivariant predictors. We investigate connections between equivariance and risk unbiasedness and show that uniquely best risk unbiased predictors are almost equivariant. We apply our theoretical results to some illustrative examples.

Key Words: Loss invariance; Maximal invariant; Minimum risk; Risk unbiased; Transformation group.

Building-type classification based on measurements of energy consumption data

Ying Ni, Christopher Engström, Anatoliy Malyarenko, Fredrik Wallin
Mälardalen University, Sweden

In this paper we apply data-mining techniques to a classification problem on actual electricity consumption data from 350 Swedish households. More specifically we use measurements of hourly electricity consumption

during one-month and fit classification models to the given data. The goal is to classify and later predict whether the building type of a specific household is an apartment or a detached house. This classification/prediction problem becomes important if one has a consumption time series for a household with unknown building type. To characterise each household, we compute from the data some selected statistical attributes and also the load profile throughout the day for that household. The most important task here is to select a good representative set of feature variables, which is solved by ranking the variable importance using technique of random forest. We then classify the data using classification tree method and linear discriminant analysis. The predictive power of the chosen classification models is plausible.
Key Words: data-mining, energy consumption data, classification of energy customers, clustering of energy customers.

New Tests of Symmetry Based on Characterizations, and their Efficiencies

Yakov Yu. Nikitin
Saint-Petersburg University, Russia and National Research University - Higher School of Economics, Russia

In this talk we present new nonparametric tests of univariate symmetry based on the characterization of symmetry in terms of order statistics. The test statistics are build on U-empirical distributions and turn out to be rather efficient in Pitman and Bahadur sense with respect to common alternatives.
Keywords: Testing of symmetry, Kolmogorov-Smirnov tests, Bahadur efficiency, shift alternative, Exact slope, Local optimality.

Maximum likelihood estimation in special forward interest rate models

Balázs Nyul
Department of Applied Mathematics and Probability Theory, Faculty of Informatics, University of Debrecen, Hungary

J. Gáll, Gy. Pap and M. V. Zuijlen described a special interest rate model which driven by a geometric spatial AR sheet ([1]) and introduce a new type of Heath-Jarrow-Morton forward interest rate model.
In this model we give the no-arbitrage criteria and we estimate parameters of the model (for example volatility) on special samples by maximum likelihood estimation.

Finally we observe the asymptotic behaviour of the maximum likelihood estimator in each cases.

Keywords: forward interest rate, HJM, volatility, maximum likelihood estimation

Reference

[1] J. Gáll, Gy. Pap and M. V. Zuijlen, Maximum likelihood estimator of the volatility of forward rates driven by geometric spatial AR sheet, Journal of Applied Mathematics 4 (2004), 293--309.

Value of a Firm With Exit and Suspension Options

Carlos Oliveira, Cláudia Nunes, Manuel Guerra
CEMAT/IST-Instituto Superior Técnico, Portugal

The valuation of a firm that has the possibility to exit to the market is often addressed in the literature of real options, especially after the pioneer work of Dixit and Pindyck. Since then, many irreversible investment problems have been widely studied in economic literature, but also in mathematical journals, due to the challenging questions that such problems raise. However, most firms have fixed costs and the model becomes more realistic if we consider that the firm has the option to suspend its production.

In real options models the companies (that produce the goods) make decisions concerning labor levels and capital investment. These decisions share three important characteristics. First, the decision is partially or completely irreversible and involves some sunk costs. Second, there is uncertainty over the future rewards from the investment. Third, there is some flexibility about the timing of the decision. One can postpone the decision to get more information about the future.

In this work we propose a new way of proving the value of a firm that is currently producing a certain product and faces the option to exit the market. Moreover, we study the value of a firm that has the option to suspend the production or to exit the market.

The problem of optimal exiting is an optimal stopping time problem that can be solved using the dynamic programming principle. This approach leads to a partial differential equation, called the Hamilton-Jacobi-Bellman equation.

This is a free-boundary problem, and therefore, we propose an approximation for the original model.

We prove the convergence of the solution of the approximated problem to the original one and finally, using the Implicit Function Theorem, we obtain this solution.

The value of a firm that has the possibility to suspend the production is also modeled in this work. In this case, the firm can suspend or return the production according to the demand. This problem can be formalized

as an optimal impulse control problem. Here for each optimal stopping time we choose an optimal control representing the decisions taken.
Key Words: Real Option, Free-boundary Problems; HJB Equations; Optimal Stopping Times

A new perspective of student allocation satisfaction in engineering courses in Portugal

Raquel Oliveira[1], A. Manuela Gonçalves[2], Rosa M. Vasconcelos[3]
1CMAT-Centre of Mathematics of University of Minho, 2Department of Mathematics and Applications, CMAT - Centre of Mathematics of University of Minho, 3Department of Textile Engineering, 2C2T - Centre for Textile Science and Technology, University of Minho, Portugal

This work aims to describe and characterize students allocation satisfaction in the Portuguese public higher education system through the point of view of the students, namely in the academic engineering programs. There is defined a complementary ratio, that allow the comparison with the ratio available by the Portuguese Education Ministry through the point of view of Institutions. Mann-Whitney and Kruskal-Wallis tests were performed. Through cluster analysis it was possible to identify and characterize the new ratio and make a comparison with the existing one.
Keywords: Students' index satisfaction, higher education, education policy, clusters analysis;
References
1. OECD, Organization for economic co-operation and Development Reviews of National policies for education: Tertiary education in Portugal.
Examiner'sReport.Availableathttp://www.dges.mctes.pt/NR/rdonlyres/8B016D34-DAAB-4B50-ADBB-25AE105AEE88/2564/Backgoundreport.pdf (2006).
2. DGES(s.d.).
http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso/Estatisticas/ (29 january 2014).
3.http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso/Estatisticas/(29 january 2014).
4. Oliveira, R., Gonçalves, A.M. and Vasconcelos, R.M. "An empirical study on the index of satisfaction of student allocation in the Portuguese undergraduate engineering courses", Proceedings 15th Applied Stochastic Models and Data Analysis (ASMDA2013), International Conference, Mataró (Barcelona), Spain, 675-681 (2013).
5. Conover, W.J. Practical Nonparametric Statistics. Third Edition, John Wiley & Sons, New York (1999).

6. Gore, P.A. Cluster analysis. Handbook of applied multivariate statistics and mathematical modeling, ed. H.E.A. Tinsley and S.D. Brown, San Diego, CA: Academic Press New York, Academic Press (2000).
7. Barnett, V. Interpreting Multivariate Data. John Wiley & Sons, Sheffield (1981)

A study of the date of death of a family in Cuba in dependence of the date of birth

Ricardo Osés Rodríguez[1], Rigoberto Fimia Duarte[2], Guillermo Saura González[1], Alfredo Pedraza martínez[1], Nancy Ruiz Cabrera[1], Julia Socarras Padrón[1]
[1]Provincial Meteorological Centre of Villa Clara, [2]Unity of Surveillance and Antivectorial Fight, Centre of Health and Epidemiology, CUBA

Presently work is modeled the possible date of the members' of a Cuban family death using the date of its birth, for this work the dates of birth and 20 members' of this family death were used, the same one presents data from 1865 up to 1996, the simple statistical regression was used for the modeling and presage of the data, as well as the methodology ROR, a model is obtained that explains 96.1 variance% with an error of 1485 16:07 for the simple regression, for the methodology ROR is explained the 100 of the cases with an error of 1794 03:53, the dependent variable of more weight it was the date of birth multiplied by 0.822 for the simple regression, the correlation among the year of death with the predicted value is of 0.961 significant to 99% for the simple regression and 0.964 for the methodology ROR. We presented a graph of the real value and predicted for each member, the statistical package SPSS Version was used 13.
Keywords: long-term forecast, deaths, time expectancy of life, Cuba, Mathematical modeling, life

Long term forecast of meteorological variables in Sancti Spiritus. CUBA.

Ricardo Osés Rodríguez[1], Rigoberto Fimia Duarte[2], Guillermo Saura González[1], Alfredo Pedraza martínez[1], Nancy Ruiz Cabrera[1], Julia Socarras Padrón[1]
[1]Provincial Meteorological Centre of Villa Clara, [2]Unity of Surveillance and Antivectorial Fight, Centre of Health and Epidemiology, CUBA

The aim of this work is aimed at modeling and forecasting with 1 year in advance a set of 7 meteorological variables, these are, as long as the wind keeps blowing over, 3m / s, 4 m / s, 5m / s , 6m / s, 7m / s, 8m / s

and 9 m / s corresponding to the meteorological station of Sancti Spiritus (Lat North 21 ° 56 ', Long 79 ° 27', Height above sea level 96.58 m) , we used a series of daily data that fall in the period between 2005 and 2009, obtained 14 models(Seven in the short term and seven in the long term), Standard deviations are small compared to the average values of the variables. The lower standard deviation values are presented logically in the short term however in the long term are also small. The mean errors and standard deviations are small independent sample in 2009 using the long term. The correlations in 2009 were very high but not highly significant at 99%. All the equations were significant at 99%. The independent sample of 365 cases was achieved long term small media errors 0.326 values for the variable in which the wind is over 9 m / s to -3.14 when the wind is above of 3 m / s. Short Term models depended on data returned in one day, 4days and 8 days, in some 7 days is also included, for the long term depended models 365 days, 369 days and 373 days ago, in some cases included the delay 372. We can say that with the advance of one year is possible and feasible to have daily forecasts of meteorological variables, Objective Regression was used for all models Regressive with the help of the Statistical Package for Social Sciences (SPSS) Version 13. The tables and graphs show the predicted and actual values for 2009. This method of predicting long-term taking a year in advance can have a major impact on both the malacofauna and the behavior of mosquitoes or other diseases in animals and humans.

Keywords: long-term forecast, wind, Cuba, Mathematical modeling, mosquitoes.

On application of gambling team technique to waiting time problems

Urszula Ostaszewska
Institute of Mathematics, University of Bialystok, Poland

Let xi_n be a sequence of i.i.d. random variables valued in a finite alphabet Ω. Consider reduced set of patterns $\mathcal{A} = A_1, A_2, \cdots A_k$ that is none of the patterns contains any other as a subpattern. Let assume that the pattern B, containing none of $\mathcal{A} = A_1, A_2, \cdots A_k$, is observed in the first m trials. We are interested in a random variable $\tau_{\mathcal{A},B}$ of the first time that one observes the pattern from $\mathcal{A}$ in the sequence xi_n after one observed the initial pattern B. Applying the gambling team technique, introduced by Li (1980) and developed by Pozdnyakov and other authors, we are able to find formulas for the expected value and generating function of the waiting time $\tau_{\mathcal{A},B}$. Moreover we consider the later waiting time of $\mathcal{A}$, i.e. the time when all patterns from $\mathcal{A}$ in the sequence xi_n are observed and waiting time till the rth occurence of

patterns from $\mathcal{A}$. Using results related with the random variable $\tau_{A,B}$ we propose the recursive formulas for the expected later waiting time and the expected waiting time of the rth occurence.

Key Words: sooner waiting time, later waiting time, gambling team technique

Vandermonde matrices, extreme points, orthogonal polynomials and moment matching

Jonas Österberg, Karl Lundengård, Sergei Silvestrov
Mälardalen University, Sweden

The Vandermonde matrix appears in many different problems, including interpolation of data, moment matching in stochastic processes with applications to computer-aided decision support and in various types of numerical analysis. Motivated by these and other applications the values of the determinant of Vandermonde matrices and some similar matrices with real elements are analyzed both visually and analytically over the unit sphere in various dimensions. The extreme points of the determinant on certain surfaces in various dimensions are identified, together with their relation to certain orthogonal polynomials and some consequences of their placement is discussed.

Key Words: Vandermonde matrix, Determinants, Extreme points, Orthogonal polynomials, Moment matching

Sensitivity Analysis of the GI/M/1 Queue with Negative Customers

Sofiane Ouazine[1], Karim Abbas[2]
[1]Department of Mathematics, University of Bejaia, [2]LAMOS, Laboratory of Modelization and Optimization of Systems, University of Bejaia, Algeria

In the area of the stochastic modelling the queueing models play a major role in performance evaluation of computer networks, communication systems and manufacturing systems. Specifically, the GI/M/1 queue is used many times if the model approximates the behavior of queues that deviate slightly from the statistical assumptions in the model. In this paper we discuss the applicability of the Taylor series approach to the numerical analysis of the GI/M/1 queue with negative customers. In other words, we use the Taylor series expansions to examine the robustness of the GI/M/1 (FIFO.∞) queueing model having RCH (Removal of Customer at the Head) to perturbations in the negative customers process (the occurrence rate of RCH). We analyze numerically the sensitivity of the entries of the stationary distribution vector of the GI/M/1 queue with negative customers to those perturbations, where we exhibit these entries as polynomial functions of the occurrence rate of RCH

parameter of the considered model. Numerical examples are sketched out to illustrate the accuracy of our approach.
Keywords: Taylor series expansion; Sensitivity analysis; GI/M/1 queue with negative customers; Numerical methods; Performance measures.

A Comparison Between Two Different Statistical Methods for Sires Evaluation in Some Milk Traits in Dairy Cattle

Elsaid Z. M. Oudah
Department of Animal Production, Faculty of Agriculture, Mansoura Universityl, Egypt

The objective of this study was to compare two statistical methods of estimating sire-transmitting ability (ETA) of three milk traits included 305-day milk yield (305-DMY), total milk yield (TMY) and lactation period (LP). These to statistical methods are Best Linear Unbiased Prediction (BLUP) and Least Squares Means (LSM). A total of 1002 lactation records of Friesian cows kept at Agricultural Research Experimental Station of Sakha. Animal Production Research Institute, Agricultural Research Center, Ministry of Agriculture, Egypt were used in the present study. Thirty-four bulls with at least five daughters each were used to compare the two statistical methods. Number of daughters per sire ranged between 5 and 87. Data were analyzed using Mixed Model Least Squares and Maximum Likelihood Computer Program of Harvey (1987). The results showed that the product moment correlations between estimates of ETA for the three milk traits estimated by the two statistical methods (BLUP and LSM) were high and positive (≥ 0.96). It could be concluded that the evaluation of Friesian cattle bulls based on BULP and LSM estimates were similar.
Keywords: Dairy cattle, Friesian, Genetic parameters, Sire evaluation, Breeding value, sire-transmitting ability, Least squares means, Best Linear Unbiased Prediction.

Effect of Statistical program on the estimates of genetic parameters of some productive and reproductive traits in Dairy Cattle

Elsaid Z. M. Oudah, Nazem Shalaby
Department of Animal Production, Faculty of Agriculture, Mansoura Universityl, Egypt

The main objective of this study was to to compare two statistical programs to estimate genetic and phenotypic parameters (heritability and genetic correlation) for productive and reproductive traits in dairy of the imported Friesian heifers and their progeny locally-born heifers under the conditions of Damietta Province located at the eastern north part of the Nile Delta in Egypt. The first statistical program was linear mixed

model least squares and maximum likelihood (LSMLMW) computer program of Harvey (1990). The second statistical program was Multivariate Derivative Free Restricted Maximum Likelihood (MTDFREML) of Boldman et al. (1995). Data were analyzed using liner mixed model least squares and maximum likelihood (LSMLMW) computer program of Harvey (1990). Two models of statistical analysis were used for analyzing factors affecting some productive traits, i.e. total milk yield (TMY, kg), 305-day milk yield (305-DMY, kg), daily milk yield (DMY, kg), Lactation period (LP, day) and dry period (DP, day). The effect of sire was used as random effect. The effect of farm, origin, season of calving and year of calving were used as fixed effects and age at first calving and days open were used as covariates. The second model was used to analyze the reproductive traits, i.e., calving interval (CI, days) and days open (DO, days). Where the same effects were used as mentioned above except days open was replaces with TMY. The same statistical models were used to estimate the genetic and phenotypic parameters for productive and reproductive traits using Multivariate Derivative Free Restricted Maximum Liklihood (MTDFREML). The Single Trait Animal Model was used to analyze the productive and reproductive traits mentioned above included individual and errors as random effects and the same fixed effects and the same covariates used in Harvey (1990) program for the same traits. Results of heritability estimates using animal model analyses according to Boldman et al. (1995) showed that all heritability estimates either productive traits or reproductive traits is very low compared with the corresponding values estimates by Harvey Program (Harvey, 1990). These differences may be due to that the animal model analyses taken into consideration the relationship between relatives in the used data which is not found in the Harvey program. This means that the software (statistical program) used in the statistical analyses may affect the values of genetic parameters.

Keywords: Dairy cattle, Friesian, Genetic parameters, heritability, productive and reproductive traits, Harvey program, Animal model

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

Latifa OURBIH-BAGHDALI, Megdouda OURBIH-TARI, Abdelnasser DAHMANI

Laboratoire de Mathématiques appliquées, Faculté de Sciences Exactes, Université de Bejaia, Ecole Nationale Supérieure des Travaux Publics(ENSTP), Algeria

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

number of loaders and trucks in the aggregate production, at a minimum cost.
A model is developed using the activity cycle diagram and the performance measures of the system are evaluated. Inefficiencies of the studied problem are identified and improvement is proposed for the quarry managers by determining the optimal strategy.
Key Words: Modelling, Activity Cycle Diagram, Discrete-Event Simulation, Monte Carlo

Survival Analysis on clinical trial in Monte Carlo simulation

Megdouda OURBIH-TARI, Mahdia AZZAL
Laboratoire de Mathématiques Appliquées, Faculté des Sciences Exactes, Université de Bejaia, Algeria

This paper focuses on the statistical comparison of Refined Descriptive Sampling (RDS) method taken from the literature for Monte Carlo simulation process and the well known Simple Random Sampling (SRS) method. For this purpose, a lifetime model whose observations are right-censored of random type is used to perform a nonparametric estimation of the survival function. A real application was conducted and its computed parameters were used as a basis for comparison. We estimate the survival function by the product limit estimation using both sampling methods and a real population. Furthermore, we computed also the bias and the variance of this estimator when entries are supposed following an exponential distribution and when entries are generated from the population distribution. The obtained results prove the efficiency of RDS over SRS in both cases.
Key Words: Nonparametric Estimation, Clinical Trial, Sampling, Monte Carlo, Simulation

MOVING AVERIGES FOR FUTURES DATA

Zidrina Pabarskaite[1], Rimvydas Simutis[1], Borisas Bursteinas[2], Aistis Raudys[3]
[1]Faculty of Electrical and Control Engineering, Kaunas University of Technology, Lithuania, [2]EMBL- European Bioinformatics Institute, UK, [3]Faculty of Mathematics and Informatics, Vilnius University, Lithuania

We analyse different methods for futures intraday time series smoothing. In trading community these methods are named moving averages and in scientific community – autoregressive models or filters. Smoothness and accuracy are two important criteria for quality evaluation. Many of research in this area concentrate on accuracy or smoothness separately. Very few looked at this as a multi criteria problem. Smoothness is important characteristic that can reduce trading costs. For practitioners

smoother moving average will allow to reduce number of whipsaw trades and accuracy to earlier detect trend and profit more. In this study we define several smoothness levels and try to find most accurate method for each smoothness level. This way we can compare various smoothing techniques like-to-like. We also propose new smoothing method where coefficients are optimised for the best smoothness / accuracy ratio. In the empirical study we use 60 most liquid futures traded in US and EU exchanges and 1000 randomly generated datasets. We examine method's performance on intraday frequencies ranging from 1 minute to daily data. New method outperforms other methods 99% on the real world out of sample data. We do not use the same data for optimisation and for evaluation. We note that the next best method is T3, that sometimes outperforms new method.

Keywords: moving average, smoothness, accuracy, weight optimization.

Youth mortality by violence in the Brazilian semiarid region

Neir Antunes Paes, Everlane Suane de Araújo da Silva
Department of Statistics of the Federal University of Paraíba, Brazil

The violence understood as a result, among other causes, of social inequalities, became not only a problem for social security, but also a public health problem of major proportions in the contemporary world, particularly for the young people in countries like Brazil. The Brazilian semiarid region is the world's largest in terms of density population and extension with 22 million inhabitants in 2010. Thus, an ecological study addressing the mortality by aggression for 137 microregions of the Brazilian semiarid, to young males, in the year 2010 was performed. The data sources used were extracted from the Mortality Information System of the Ministry of Health, the Brazilian Institute of Geography and Statistics, and the United Nations Program for Development. Two indicators were calculated for each microregion: standardized mortality rates by violence and an indicator named Reducible Gaps of Mortality, equal to attributable risk. We investigated the correlation between standardized mortality rates and a set of 154 indicators that express living conditions. 18 of them were considered as significant. By means of the multivariate technique - Principal Component Analysis - the construction of a Synthetic Indicator was performed, which was categorized in four strata reflecting different living conditions. The results showed that microregions with high values of mortality rates by agressions were present in all strata, thus contradicting some studies linking high rates of mortality due to aggression to low condition of life. The study allowed us to raise issues, and generate a database to support managers to identify the most vulnerable regions. Thus contribute to the decision making process to support the development

and implementation of more efficient rational public policies to combat mortality by violence of the Brazilian semiarid population.
Key Words: Mortality by violence, Condition of life, Young violence, Brazilian semiarid region, Multivariate Analysis.

Cointegration pairs trading strategy on derivatives

Pak-Kuen Philip Lee, Ngai-Hang Chan, Lai-Fan Pun
Department of Statistics, The Chinese University of Hong Kong, Hong Kong

The notion of cointegration has been widely used in finance and econometrics, in particular in constructing statistical arbitrage strategy in the stock and the currency markets. In this paper, the arbitrage trading strategy for derivatives based on cointegration is generalized to account for the volatility factor. Pairs of short dated at-the-money straddles of European options with positive net carry (i.e. theta) are considered, so as to capture the mean reversion property of the linear combinations of implied volatilities. A performance analysis is conducted based on historical data, suggesting that the strategy works well during low volatility regimes but poorly during high volatility regimes. To improve the performance of the strategy, criteria are imposed on the trade selection process so as to take the forecasting realized volatilities into account. Empirical study for the renewed strategy is implemented u! sing a 3-year historical data of currency options.
Key Words: Cointegration, trading strategy, implied volatility, forecasting realized volatility, ARFIMA model

Discrete semi Markov patient pathways through hospital care via Markov modelling

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

In the present paper, we 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). In terms of semi-Markov modelling we can regard each patient pathway as a state of the semi-Markov model, therefore the holding time distribution of the ith state of the semi-Markov process is equivalent to the 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 the present paper a method is applied according to which we can describe first and second moments of numbers in each semi Markov patient pathway at any time via Markov modelling. Such values are useful for future capacity planning of patient demand on stretched hospital resources. The above results are illustrated numerically with healthcare data.

Job Satisfaction of European graduates. An analysis and multiple group comparison

Ioulia Papageorgiou[1], Aglaia Kalamatianou[2]
[1]Department of Statistics, Athens University of Economics and Business,
[2]Department of Sociology, Panteion University of Social and Political Science, Greece

Undoubtedly, job satisfaction is one of the most important aspects of work with significant implications in business, labor relations and to an individual's life in general. Its study has origin the beginning of the last century and has attracted the interest of researchers from different disciplines. However, due to the difficulty, complexity and the convoluted of the concept, -fact for which there is general agreement-, there is no yet a single conceptual model that perfectly explains the issue. An important question which also arises is how to assess differences of Job satisfaction among different strata of employees. In this paper, using data from a large-scale European survey concerning employment and educational aspects of graduates from different countries, we combine explanatory and confirmatory factor analysis methods as well as logistic regression, to assess the common model that best explains job satisfaction of the total sample of graduates. We also investigate possible differences of implementing the model to different groups formed taking into account independent variables like graduates' country of origin, gender, level of education and field of studies.

A study of an interval scale for a motivation test

Artur Parreira[1], Ana Lorga da Silva[2]
[1]CPES-ECEO, ULHT, Portugal and CESGRANRIO, Brasil, [2]CPES-ECEO, ULHT, Portugal and CEDRIC – CNAM, France

This study deals with the development of interval scales, which combine the qualitative dimension of language with the corresponding quantitative dimension. The qualitative dimension is represented by adverbs of quantity and frequency, currently used by people, when they assign a value to things; the quantitative dimension was determined by an empirical study of the quantitative value of quantity and frequency adverbs.

This article is focused on adverbs of quantity.
A multivariate analysis is made, mainly using Classification methods (Gordon, 1999) and Principal Component Analysis (Jollife, 2002).
The interval scale was constructed to be used in a test of motivational profiles (Pinder, 2008) theoretically based in the paradigm of complex behavior processes (Parreira, 2006).
Keywords: Motivational profiles; Interval scales; Adverbs of quantity; Classification methods, PCA

References

1. Gordon, A. D. (1999), Classification, Chapman & Hall, London.
2. Jollife, I.T. (2002), Principal Component Analysis, Springer, New York.
3. Parreira, A. (2006), Gestão do Stress e da Qualidade de Vida, Monitor, Lisboa.
4. Pinder, C. (2008), Work motivation in organizational behavior (2nd ed.). Psychology Press, New York US.

An Optimal Band is using for predicted the buy and sell signals of stocks

Parmod K. Paul, Vivek Vijay
COE-System Science, Indian Institute of Technology Jodhpur, India

In this paper, we propose a new trading band, namely the Optimal Band. This is useful to see buy and sell signals. This band uses a linear function of local and absolute extreme values. The function is optimized by using solver in MS Excel. This band has a property of including 100 % data points from a given financial time series. We compare the patterns generated by optimal band and that of by Bollinger Band for BSE Sensex April 2007 to Oct 2013 and Crude oil intraday data 22 march 2007 10:00 to 15:45 at minutes frequency.
Keywords: Optimal Band, Bollinger Band, Buy and sell signals, Linear function

Is it worth using a fuzzy controller to adjust the mutation probability in a genetic algorithm when the input variable is the number of iterations?

André **G. C. Pereira[1] , Enrico Colossimo[2], Bernardo B. de Andrade[3]**
[1]Mathematics Department, Universidade Federal do Rio Grande do Norte, Natal/RN, [2]Statistics Department - Universidade Federal de Minas Gerais, Belo Horizonte/MG, [3]Statistics Department - Universidade Federal do Rio Grande do Norte, Natal/RN, Brasil

In recent years, several attempts to improve the efficiency of the Canonical Genetic Algorithm have been presented. The advantage of

the elitist non-homogeneous genetic algorithm is that variation of the mutation probabilities permits the algorithm to broaden its search space at the start and restrict it later on, however the way in which the mutation probabilities vary is defined before the algorithm is initiated. To solve this problem various types of controllers can be used to adjust such changes. This work presents a study that if the input variable is the number of iterations of the algorithm then the usage of a fuzzy controller to adjust the mutation probability of an elitist non-homogeneous genetic algorithm produce a high cost and it can be substituted by a stochastic controller without loss of performance. And finally, numeric simulations are used to illustrate that the performance of the stochastic controller is statistically as good as that of the fuzzy controller in all examples performed.
Keywords: Stochastic controller, Fuzzy controller, Genetic Algorithms.

Insuring wind farm from energy underproduction: a semi-Markov approach

Filippo Petroni, Flavio Prattico, Guglielmo D'Amico
University of Cagliari, Italy

In this paper we present and analyze the problem of energy production from a wind farm. In particular we focus our attention on the stochastic nature of wind speed and on possible insurance product to hedge a wind farm from underproduction. We propose an insurance product (swap) that is able to solve this issue by using an indexed semi-Markov process for wind speed.
Key Words: Stochastic Modeling, Wind Speed, Swap, Wind Farm

Modelling of the multivariate skewness measure distribution

Margus Pihlak
Tallinn University of Technology, Estonia

The problem of several presentation concerns the estimation of statistical models. This is the problem of skewness or lack of symmetry. That means the distribution of statistical models residuals are frequently non-gaussian as Kolmogorov-Smirnov test shows. In this case the skewness has to be estimated for testing the goodness of models. The confidence intervals of that parameter have to be found. This enables us to improve the diagnosis of statistical models. By means of skewness confidence intervals we can estimate the influence of outliers. These outliers are typical in forestry. The main question is: does zero value belong to the estimated confidence interval.

Firstly we present some useful results in multivariate statistical analyses. In this presentation we refer mainly the papers K.V. Mardia (1970), M. Pihlak (2004) and T. Kollo (2008). Then we generalize the central limit theorem to the multivariate skewness measure. Finally we apply the obtained results on statistical models diagnostics.

Keywords: Multivariate skewness measure, Skewness measure distribution, Analyses of statistical models residuals

References

1. Kollo, T. (2008) Multivariate skewness and kurtosis measures with an application in ICA. Journal of Multivariate Analyses, 99, 2328-2338.
2. Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis measures with applications. Biometrika, 57, 519-530.
3. Pihlak, M. (2004) Matrix integral. Linear Algebra and its Applications. 388, 315-325

The use of meta-analytic integration of chemical and biological data in Agroforestry Systems in Brazil

Maria de Lourdes Pinheiro Ruivo, Adrane Rocha
Museu Paraense Emilio Goeldi, Brazil

This work seeks to organize a database in order to later metaálise chemical and biological soil under agroforestry (SAF) in Brazil data. So, we organized a general search for articles by SAF's, and groups them based on the following characteristics: location, crop type and management, depth of the organic layer. The databases will be cosntruida through search in libraries and search sites in the CAPES portal strategy. The main ones are: Web of Science, Science Direct, Wiley Blackwell (Multidisciplinary), SciFinder Scholar, Springer - (Agricultural Sciences and Earth Sciences), Scopus (Earth Science) and SciELO: Covers a collection of Brazilian scientific journals. The search terms are: agro - systems (isolated); - agroforestry systems (combined): Sustainability, soil quality, family farms, organic matter, microbiological indicators, chemical indicators, soil fauna, microbial biomass, litter input, rainfall, temperature, seasonality; management; cycling of nutrient cycling, soil fertility, grain size, texture. The limitations of search are adopted: Geographic Range: Works where areas of study are in Brazil; Period: 2000-2012, only selected papers will be published starting from the top - "include published studies that were reviewed" But as significant results are published more than insignificant then: There is no significant articles on the understanding that it would be a way to bias the choice of certain articles qualis deleted. By applying the techniques of meta- analysis is expected to define a set of indicators of soil quality under SAF'sea indication of more crops suitable to this system of management

Nonlinear reconstruction of signals by diffusion maps

Lúcia M. S. Pinto; Ricardo Fabbri; Francisco D. Moura Neto
Rio de Janeiro State University, Brazil

Diffusion maps have shown very useful for dimensionality reduction of data sets of high dimension. This method has been introduced by Coifman et al. [1]. Following the program set up by Munford and Desolneux [2], which calls for a feedback architecture for data recognition and reconstruction, we construct a procedure for regularized reconstruction of signals, based on the distance from the training data and the optimization from diffused data. These ideas have also been applied to dimensionality reduction by the classic principal components analysis, however producing a nonlinear reconstruction of the signal, allowing for the treatment of nonlinear data. The results show the method is robust.

Key Words: Diffusion maps; reconstruction of signals; dimensionality reduction

Reference

[1] R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker,"Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps," Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 21, p. 7426, 2005.

[2] D. Mumford and A. Desolneux, Pattern theory: the stochastic analysis of real-world patterns. Natick, MA: A K Peters, Ltd, 2010.

Economic and financial viability of peri-urban dairy farms in Mexico's central Valley: 2010-2018

Rodolfo R. Posadas-Domínguez[1], Nicolás Callejas-Juárez[2], Carlos Arriaga-Jordán[1], Francisco E. Martínez-Castañeda[1]
[1]ICAR, Universidad Autónoma del Estado de México, [2]FZE, Universidad Autónoma de Chihuahua, México

The effects on the economic and financial viability when different risk scenarios are included in representative production units (PRU) of small-scale milk were determined using Monte Carlo simulation. Scenarios were: 1) Real; and 2) the purchase of 100 % of inputs for livestock feed in production process. Panel Technique was use and 22 dairy producers with similar levels of technology, production scale, geographic location and product market were selected for de PRU, also 4 disciplinary specialists and a representative of the regional dairy association where included. The size of the PRU was 8 cows on production (PRU8). This

PRU represents 130 producers in the area. The economic and financial projections of micro and macroeconomic variables estimated by the Secretary of Agriculture, Livestock, Rural Development, Fisheries and Food (SAGARPA) were included to generate the stochastic components of the models analyzed. The econometric program Simetar© was used to quantify the risk and predict the economic and financial impacts on the planning horizon from 2010 to 2018. The results indicated a viable economic and financial situation in real scenario but not at Scenario 2. The probability of economic success as measured by the rate of return on assets was 80% for scenario 1. On the contrary, Scenario 2 provided a probability of disinvestment of 40%, which is compounded from 2014 until the end of the planning horizon. The probability of financial success mediated the net present value (NPV) was 93% for scenario 1. Beginning 2014, the financial unsustainability of scenario 2 is sharpened and the NPV becomes negative.
Keywords: Prospective analysis, livestock, stochastic models.

The recent trends in mortality in Poland

Alina Potrykowska
The Government Population Council of Poland, The Polish Academy of Sciences, Poland

The health policy should, contribute to the implementation of the population objectives in Poland, as it has reached a critical stage of its demographic development. For a couple of years, the number of residents has not been growing, which results from the generally zero, and in some years even negative, balance between the number of births and deaths. At the same time, for almost the entire post-war period, Poland has been losing its population for the benefit of other countries, especially Western European and the USA.

These unfavourable demographic development features regarding births and deaths are also largely related to the level and quality of public health. The most significant disadvantages displayed by Poland in various areas of public health, as compared to Western European countries are: The average life expectancy of women in Poland is still by 8.5 years longer than the average life expectancy of men, and this gap is much bigger than in the countries with the lowest mortality rate. Despite the low level of deaths among newborn children, it is still twice as high as in the European countries which are leading in this respect.

The level of premature deaths, i.e. before the age of 65 years, is still by 50% higher than in the “old” EU countries. This is related to the level of mortality due to circular diseases, cancer or the so-called extraordinary deaths, which in Poland is still higher than in the reference countries.

This paper presents and analysis of the cancer incidence and mortality in Poland. It has been supplemented by a short-term forecast of incidents and mortality for the most common tumor locations.
Key Words: Health state, mortality, Poland

Parameter estimation of a particular class of 2n dimension Ornstein-Uhlenbeck processes

Ana Filipa Prior[1], Marina Kleptsyna[2] and Paula Milheiro de Oliveira[3]
[1]*High Engineering Institute of Lisbon and CMUP, Lisbon, PORTUGAL,*
[2]*Laboratoire Manceau de Mathématiques Université du Maine, France,*
[3]*Engineering Faculty of the University of Porto and CMUP, PORTUGAL*

We consider a 2n dimension OU process for which the diffusion matrix is singular. This process is used as a model for the dynamic behavior of vibrating engineering structures as bridges or suspended roofs, among others. We study the problem of estimating the vibrating frequencies of the structure or, equivalently, the parameters of the SDE that governs the OU process. The maximum likelihood estimator of the drift matrix derived by Koncz[3] is revisited, as well as the way properties of the estimator are established. The local asymptotic normality of the estimator is analyzed in detail. Since general regularity conditions do not hold in this case (the diffusion matrix is singular), theoretical results from the classic literature on the subject do not immediately apply and we recall both classical theorems (see Prakasa Rao[4]) as well as an alternative approach based on the Laplace transform. Simulation results are presented to illustrate the convergence to a normal distribution.
Acknowledgments Research funded by the European Regional Development Fund and by the Portuguese Government through the FCT under projects PEst-C/MAT/UI0144/2011 and PROTEC/IPL
Keywords: Stochastic differential equation, maximum likelihood estimator, Ornstein-Uhlenbeck process.
References
1. Arato, M. (1982), Linear Stochastic Systems with constant coefficients. A statistical approach., Lectures Notes in Control and Information Sciences, 45. Springer-Verlag, Berlin.
2. Basak, G., and Lee, P., Asymptotic properties of an estimator of the drift coefficients of multidimensional Ornstein-Uhlenbeck processes that are not necessarily stable., Electronic Journal of Statistics, Vol.2, 1309-1344 (2008).
3. Koncz, K., On the parameter estimation of diffusional type processes with constant coefficients (elementary Gaussian processes)., Journal of Anal. Math. 13, No. 1, 75-91 (1987).

4. Prakasa Rao, B.L.S. (1999). Statistical Inference for diffusion type process. Kendall´s Library of Statistics, 8. Edward Arnold, London. Oxford University Press, New York.
5. N. Lin, S. V. Lototsky. Undamped harmonic oscillator driven by additive Gaussian white noise: A statistical analysis, Communications on Stochastic Analysis, 5, No. 1, pp.233-250 (2011).

Control Charts for Arbitrage-Based Trading

Stelios Psarakis[1], Kostas Triantafyllopoulos[2], Aggeliki Vyniou[1]
[1]Department of Statistics, Athens University of Economics & Business, Greece,
[2]School of Mathematics and Statistics, University of Sheffield, UK

This talk concerns the role and contribution of SPC methods to pairs trading, a relative-value statistical arbitrage trading technique. Pairs trading and its numerous extensions have gained increased attention over the recent years and is a popular trading strategy among hedge funds and investments boutiques. The core idea of pairs trading is “buy low” and “short-sell high” and it is based on the assumption that the low-valued asset will gain value and the high-valued asset will lose value, so that the two assets are co-evolving or mean-reverting. Several authors have recently suggested that this mean-reversion may be appropriate only at some periods of time (in which profits may be realized) while in other periods of time mean-reversion is lost (resulting to significant loss, e.g. you may buy an asset which loses its value). In this talk we propose the use of appropriate SPC methods in order to detect mean-reversion, hence to identify tradable periods. The literature on pairs trading will be reviewed and examples will illustrate the need for a monitoring procedure. Control charts for bivariate autocorrelated processes are proposed for the detection of mean-reversion. Throughout the talk the methods will be illustrated with closing prices of suitable pairs.
Keywords: Bivariate control charts, Autocorrelated process, Financial Processes.

Models for the Demographics of Commercially-Utilized Fish Populations

Terrance J. Quinn II
Juneau Center, School of Fisheries and Ocean Sciences, University of Alaska Fairbanks, USA

Models for the demography of commercially-utilized fish populations, so-called “fishery models”, attempt to provide the same type of information as models for human populations, but are fundamentally different in that

fish populations are not observed directly, as in a census. There are several approaches for the estimation of population parameters: (1) using sampling methods such as mark-recapture, distance, and catch-effort, (2) using simple population models based on exponential and logistic growth, and (3) more complex models that incorporate biological processes such as reproduction, mortality, and growth. These models are surveyed in Quinn and Deriso (1999, Quantitative Fish Dynamics, Oxford). Contemporary models integrate diverse data sources such as indices of abundance from the fishery(ies) and from biological surveys, sampling information about age, length, and sex composition from the fishery and from biological surveys, estimates of reproductive parameters, such as maturity and fecundity, and other life history information (e.g., natural mortality, predation, and disease). Models are necessary to effect the integration. There are three major components of these fishery models: (1) the population dynamics in terms of equation to describe the process of year-classes over time, variation in terms of statistical errors in measurement (sampling) and process (population parameters as random variables), and (3) an objective function for minimizing the differences between observations and model predictions as a function of population parameters. I review these contemporary models and provide an overview of recent developments to incorporate temporal and spatial variation through stochastic, state-space, and Bayesian models.

Control Charts for Zero-Inflated Processes with Estimated Parameters

Athanasios C. Rakitzis, Philippe Castagliola
LUNAM Universite, Universite de Nantes & IRCCyN UMR CNRS 6597, University of Nantes, France

Zero-inflated probability models are used to model count data that has an excessive number of zero counts. These models are mostly useful in modeling high-yield processes that produce a low fraction of non-conforming units or health-related processes where it is of interest the monitoring of a rare disease. Shewhart-type control charts have been proposed for the monitoring of zero-inflated processes. Usually their performance is evaluated under the assumption of known process parameters. However, in practice their values are rarely known and have to be estimated from an in-control historical Phase I data set. In this work, we investigate the performance of Shewhart-type control charts for zero-inflated Poisson (ZIP) and zero-inflated binomial (ZIB) processes when the process parameters are estimated from a Phase I data set of size m. Design aspects concerning the necessary size m of the Phase I

data set, so as the in-control performance of the examined charts is the desired one, are also discussed.
Keywords: Average run length, Moment estimator, Probability generating function, Standard deviation run length, Shewhart control charts, Zero-Inflated binomial distribution, Zero-Inflated Poisson distribution.

Gini coefficient for beta distributed scores

Martin Řezáč
Department of Mathematics and statistics, Faculty of Science, Masaryk University, Czech Republic

It was justified in the literature that the Beta distribution provide a reasonable approximation of distribution of a credit scores, which are the outcome of credit scoring models, i.e. models which are used to determine the probability of default (i.e. when the client fails to meet his or her credit obligations). Credit scoring models are used in practice in the majority of decisions related to the granting of credits, and are thus inherently part of the majority of processes (approval, enforcement, commercial, etc.) in the financial sector.
Besides Kolmogorov-Smirnov statistic, AUC or J-divergence (also called Information value) Gini coefficient is widely used to assess discriminatory power of credit scoring models. Moreover, Gini coefficient is one of the most used statistics to assess discriminatory power of classification models in many other fields.
Empirical estimator based on ROC using trapezoidal rule with deciles of scores, which is one of common way how to compute the Gini coefficient, may lead to strongly biased results. Nevertheless, it can be shown that estimator based on Wilcoxon-Mann-Whitney statistic is unbiased. However, the disadvantage of this estimator is its computational demands.
The main objective of the contribution lies in providing a simulation study showing properties of both parametric and non-parametric estimators of Gini coefficient for credit scoring models with Beta distributed scores. Specifically, it is focused on the bias and mean squared error (MSE) of the estimators according to size of data sample and selected parameters of given data sample. Furthermore, the formula for Gini coefficient for beta distributed scores is derived and its computational issues are discussed.
Keywords: Gini coefficient, Bias, MSE, Beta distribution, Credit scoring.

Stochastic Evolution of New York Stock Market Distributions

Paulo Rocha[1], João P. da Cruz[2], Frank Raischel[3], Pedro G. Lind[4]
[1]Mathematical Department, FCUL, University of Lisbon, Portugal, [2]Closer Consultoria Lda, Portugal, [3]Instituto Dom Luiz, CGUL, University of Lisbon, Portugal, [4]ForWind and Institute of Physics, University of Oldenburg, Germany

Using data from the New York stock market, extracted from the Yahoo platform (http://finance.yahoo.com) every 10 minutes since January 2011, we test four different bi-parametric models to fit the correspondent volume-price distributions at each 10-minute lag: the Gamma distribution, the inverse Gamma distribution, the Weibull distribution and the lognormal distribution. In each case, the value of the pair of parameters is recorded, composing a bivariate time-series, which is then analyzed as a stochastic process. Assuming that the evolution of the two parameters is governed by a two-dimensional coupled Langevin equation, we derive the corresponding drift vector and diffusion matrix, which can then provide physical insight for understanding the mechanisms underlying the evolution of the stock market.
Keywords: Stochastic Data Analysis, Stock Market.

How Long It Takes To Get Social Insurance? Informality Dynamics on the Egyptian Labor Market during the 1998-2012 Period

Rania Roushdy[1], Irene Selwaness[1,2]
[1]Poverty Gender and Youth Program, Population Council, Cairo office, [2]Faculty of Economics and Political Science, Cairo University, Egypt

This paper aims at better understanding the extent of social security coverage on the labor market of – one of Middle East countries - Egypt. Two questions pertaining to social insurance coverage in the private sector are tackled: Who has access to social insurance (SI) coverage, and how long it takes to get such access? Using three rich nationally representative Egyptian labor market surveys conducted in 1998, 2006, and 2012, the analysis will first estimate the individual- and firm-level determinants of being enrolled in the SI system, in order to answer the "Who" question. Second, A hazard model will be used to address the "how long" question by estimating the duration needed, since first entry to the labor market, to get covered by the SI system. Preliminary findings show men, older, married, better educated and white collar highly skilled wage workers are more likely to have social insurance coverage.

Transition from School-to-Work in Egypt: An update on Young People Labor Market Conditions in the Wake of the January 25th Revolution

Rania Roushdy[1], Irene Selwaness[1,2]
[1]*Poverty Gender and Youth Program, Population Council, Cairo office,*
[2]*Faculty of Economics and Political Science, Cairo University, Egypt*
rroushdy@popcouncil.org, irene.selwaness@feps.edu.eg

Introduction

Egypt was at a stage in its demographic transition with a marked 'youth bulge', a period in which the proportion of youth in the population increases significantly compared to other age groups. This youth bulge has always posed opportunities as well as challenges for development. While it represents a "demographic window of opportunity" for economic growth (World Bank 2007), this demographic bonus is not automatic; it depends on the country's social and economic policy responses in the areas of youth education, employment and health. Such growing youth population in Egypt was more of a challenge, since it led to severe labor supply pressures on the Egyptian labor market. During the last two decades, the prevalence of long youth unemployment phases and informal jobs does not only testify the hardship position young people were facing, but also leads to doubts whether Egypt will reap the demographic bonus harvest or not.

The hard labor market conditions following the revolution have been widely documented in the news, but very little data have been available to appropriately analyze those effects. The Egyptian largest demographic cohort (the young people) is currently making its way to adulthood. Such transition to adulthood along with the slowdown of Egypt's economy following the revolution, particularly in the tourism and services sectors, mandates a detailed analysis of the impact of the revolution on the employment and unemployment situation of young people in Egypt (Assaad and Kraft 2013). This is the main objective of this paper. The paper will update the state of knowledge on young people school-towork transition in Egypt in the wake of the January 25th revolution. It will investigate how their transition from school to first job was affected from before to after the revolution as well as how the revolution has affected their job quality, formality status and their job market mobility.

Data and Methods

The analysis of this paper relies on panel data from the two nationally-representative Surveys of Young People in Egypt (SYPEs): SYPE 2009 and SYPE 2013. SYPE 2013 is a follow-up survey to the Survey of Young People in Egypt 2009, re-interviewing the same sample of young people surveyed in 2009 in all governorates of Egypt expect the Frontiers governorates. The initial 2009 round targeted a

nationallyrepresentative sample of 15,029 young people aged 10 to 29, thus encompassing both "youth" and "adolescents." The purpose of this age range was to track young people throughout the complete duration of their transition to adulthood, allowing for an extended period to account for the phenomenon of delayed marriage and in some cases transition to productive work (Assaad and Barsoum 2007). In the 2013 round we will go back to interview about 13, 800 young people who should now age 14 to 33.

The fact that SYPE was fielded less than two years prior to the revolution also offers a unique opportunity to capitalize on the natural experiment of the January 25th revolution and gain a before-and-after picture of the economic, social and political situation of young people during this critical period of Egypt's history. SYPE is the most comprehensive source of data on young people in Egypt today. Both surveys are rich sources of information on adolescents and youth conditions in Egypt before and after the revolution, including education, employment status, unemployment, job mobility, wage earnings, migration, family formation, health and sexuality, and civic and political participation. They also contain great deal of information on the household members' demographic and socioeconomic characteristics, housing conditions, ownership of durable goods, and access to basic services and infrastructure.

Using data from those two rich nationally representative surveys will enable us to address the main questions of this study. First, we will estimate the duration young people endure to find their first job as well as the age at first entry in 2009 and in 2013 using non-parametric (Kaplan Meier failure estimates) and parametric methods (hazard models). The comparison of the estimated duration to find the first job and the age at first job will shed light on how the transition from school-to-work was affected after the revolution. Second, the main determinants of employment status and school-to-work transition will be estimated, separately for young men and women, using multinomial logit models for the employment status and probit models for the school-to-work transition. This will allow discovering any changes that happened in the determinants of employment status, its structure and its formality status before and after the revolution. Third, for young people whose transition from school-to-work was successful and relying only on SYPE2013, the probability of moving to a formal job is estimated distinguishing between before and after the revolution. Besides controlling for individual-and household-level and job characteristics, the effect of the revolution will be captured by introducing a dummy for the year 2011-2012. This should enable us to analyze the impact of the revolution on the mobility pattern of young people from informal to formal employment.

Preliminary/Expected Findings

Our preliminary analysis so far sheds light on the hardship position which young people have been suffering from during the 2009-2013 period. In

particular, young women seem to be the category that bears most of the brunt of the economic slowdown due to the revolution. They faces longer unemployment spells as well as to be leaving the labor force as the youth bulge ages. Young people first jobs are now increasingly in irregular wage work instead of informal one. Such evolution away from informal wage work shows more precarious work conditions that the young generations are facing.

References

Assaad, Ragui and Ghada Barsoum. 2007. Youth Exclusion in Egypt: In Search of "Second Chances." Middle East Youth Initiative Working Paper, No.2. Washington, D.C.

Assaad, Ragui and Caroline Kraft. 2013. "The Evolution of Labor Supply and Unemployment in the Egyptian Economy: 1988-2012." Economic Research Forum Workshop on Preliminary Results of Egypt Labor Market Panel Survey 2012 June 9th, 2013, Cairo, Egypt

World Bank. 2007. World Development Report 2007: Development and the Next Generation. Washington, DC: World Bank.

Social vector of mortality due to cardiovascular diseases in Russia

Tamara P. Sabgayda, Victoria G. Semyonova
Federal Research Institute for Health Organization and Informatics, Russia

We have analyzed the dynamics and structure mortality from cardiovascular diseases (CVD) in Russia in comparison with the situation in European countries. The comparative analysis was done for dynamics and structure of mortality from CVD of the Russian population in different types of communities in the context of the village - the city - the capital.

Key Words: Mortality from cardiovascular diseases, Mortality structure, Inequality in access to health care, The implementation of federal programs to reduce cardiovascular mortality, Administrative resource in reducing mortality

The Joint Distribution of Sums of Success and Failure Runs in a Sequence of Bernoulli Trials and Its Use for Discriminating Membrane Proteins

Athanasios Sachlas[1], Pandelis Bagos[2], Sotitis Bersimis[1]
[1]Department of Statistics and Insurance Science, University of Piraeus,
[2]Department of Biomedical Informatics, University of Thessaly, Greece

Discriminating integral membrane proteins from water-soluble ones, has been over the past decades an important goal for computational

molecular biology. A major drawback of methods appeared in the literature, is the fact that most of the authors try to solve the problem using computational techniques. Specifically, most of the proposed methods require an appropriate data set for training, which means that the results depend heavily on the suitability of the data set, itself. Motivated by these facts, in this work we develop a formal discrimination procedure, which is based on the exact distribution of a two dimensional runs-related statistic defined on the sequence of hydrophobic and polar residues along the sequence. Specifically, for setting up our discrimination procedure, we study thoroughly the exact distribution of a bivariate random variable which accumulates the exact lengths of both success and failure runs of specific lengths. In order, to investigate the properties of this bivariate random variable, we use the Markov chain embedding technique.
Finally, we apply the new procedure to a well-defined data set.
Keywords: Runs; Scans; Patterns; DNA Analysis, Proteins Analysis; Bioinformatics; Markov Chain Embeddable Random Variables; Bivariate Markov Embedded Random Variables of Polynomial Type; BMVP

Fractional Stable Distribution in Gene Expression

Viacheslav V. Saenko
Ulyanovsk State University, Russia

Statistical investigations of gene expression had been carried out in the last two decades reveal some universal properties which their inherent. First of them it is power law behavior of gene expression profile at big values of expression. Such behavior inherent for gene expression for all organisms from bacteria to human tissues [1]. For description these profiles was used the Zipf's law [2], the lognormal laws [3] and others.
In this work for description gene expression the fractional stable distribution was used. The fractional stable distribution is limit distribution of sums identical distributed random variables. The probability density function is expressed through Melin's transformation of two Levy stable distribution

$$q(x;\alpha,\beta,0)=\int_0^{\infty} g\left(xt^{\beta/\alpha};\alpha,0\right)t^{\beta/\alpha}g(t;\beta,1)dt.$$

According to the fact that gene expression profiles have power law asymptotic, as the fractional stable distributions, we propose that gene expression profiles can be described by fractional stable distributions.
Various tissues were processed, for example adult rat, Male Gametophyte Development in Rice, C. elegans cells, Drosophila S2 cells, clinical S. aureus strains and others. The parameters of fractional stable distributions were statistically estimated by experimental data. It was obtained well agreement between empirical distribution and

fractional stable distribution for the parameter which have been estimated for processed tissues.
In the work the fractional stable distribution and algorithm of statistical estimation of the parameters are described. The results of comparisons of theoretical and empirical distributions are presented.
The work was done by support Ministry Education and Science of Russian Federation (agreement No. 2014/296) and Russian Foundation for Basic Research (grants No. 01-12-00660)

Analysis survival with Aalen and Cox Models For delaying blindness in patients with diabetic

Alir Reza Safaei, Seyyed Ali Mirebrahimi
Statistical center of Iran, Iran

Regression models use more often in carcinogenesis studies .These models are used for estimation of the effect of covariates on survival functions [1].In this paper we give example to illustrate Aalen model that uses in additive hazard models. See -David W Homser [1999] Nikulin, Hong-Da Isaac Wu [2006].some regression models were used in Jerald F.Lawless [2002], -T.Martinussen and T.Sheike [1999].
Keywords: additive hazard model, hazard rate, proportional hazard

A Stochastic Modelling of The Extraction Process: Response Surface Methodology

Selin SAHIN[1], Ruya SAMLI[2]
[1]Department of Chemical Engineering, Istanbul University, [2]Department of Computer Engineering, Istanbul University, Turkey

In this study, polyphenolic material amount in olive tree (*Olea europaea*) leaves extracted by ultrasound-assisted extraction (UAE) were investigated. The effect of three parameters such as temperature (25–35–45 °C), time (60–90–120 min) and pH (3–7–11) were determined by both experimental and Response Surface Methodology (RSM) techniques. While applying RSM to the experimental results, a computer simulation was written in MATLAB simulation program to see the responses for each parameter scenario. The experimental results were compared with those of calculated from RSM. The differences between experimental and calculated results were expressed as percentage (% Dif). RSM was found to be an appropriate method for the UAE of olive leaves, owing to its high accuracy.

Keywords: Simulation; Process parameters; Response Surface Methodology; Modelling; Olive tree (*Olea europaea*) leaves; Ultrasound-assisted extraction.

Preservation of ageing classes in deterioration models with independent increments

C. Sanguesa, F. G. Badía, J.H. Cha
Department of statistical methods, University of Zarangosa, Spain

In the present work we consider ageing properties in a deterioration model in which the stochastic process measuring deterioration is a process with independent increments. Preservation of Increasing and Decreasing failure rates (IFR and DFR), as well as Decreasing Reversed Hazard Rate (DRHR) are considered. We also take into account the preservation of log concave and log convex densities. Our main results are based on technical results concerning preservation of log concave and log convex functions by positive linear operators, and include the study of stochastic ordering properties among the random variables in the process.
Key Words: wear process, ageing class, stochastic order, log concavity (log convexity)

The Effect of Using Independent Increments Devoid of Path Regularity Properties in Process Estimation

Lino Sant
Department of Statistics and Operations research, Facutly of Science, University of Malta, Malta

Simulated data used in the testing phase of the estimation of stochastic processes with independent, stationary increments like Lévy processes, is many times generated through independent sampling from the relevant distribution. However cumulative series thus obtained cannot be treated naively as though it comes from discretization of a randomly selected path of some standard type of process of say the Lévy type. Regularity properties of the paths are completely ignored in such exercises. Testing of estimators within these contexts could consequently be easily invalidated or at least be subject to statistical flaws. In this paper we study by how much does this neglect of sample path properties affect the estimation process in some specific cases.
Keywords: stationary independent increments, simulated paths, regularity of paths

Estimation of Lévy process through Stochastic programming

Lino Sant, Mark Anthony Caruana
Department of Statistics and Operations research, Facutly of Science, University of Malta

Estimation of Lévy processes with the use of the characteristic function has lately shifted much of its attention to nonparametric settings. However the parametric context still offers scope for study. The nature of neighborhoods of the minima sought for by the integrated square error estimator (ISEE), and its variants, could be meaningfully related to a number of useful properties possessed by the estimator. Furthermore the numerical problems associated with the actual computation of parameter estimates have not been given exhaustive attention. In this paper through a slight reformulation of the ISEE formula, local geometric features of the optimal solution used in ISEE are studied. This formulation is subsequently proposed within a stochastic programming framework. The latter provides a powerful, productive methodology and an alternative theoretical framework which are entertained within this study. Results are presented and discussed.
Keywords: Lévy Processes, Characteristic Function, Stochastic Programming.

ML estimates of Modified Gravity Model with Skew-Normally Distributed Errors

Diana Santalova
Institute of Mathematical Statistics, University of Tartu, Estonia

Many models have been suggested for passenger and migration flows estimation, and special attention has been paid to the gravity models and their modifications. A modified gravity model was considered by A. Andronov and D. Santalova in [1, 3]. The model estimation under the assumption that input data and the error term are distributed normally is carried out in [1, 3]. Although non-symmetry of the passenger flows may violate the assumption about normality of error distribution. Two generalizations of the model have been suggested (with multiplicative and additive skew-normally distributed error term), and a method of their estimation was proposed in [4]. However, some of the obtained ML estimates of the models parameters can be biased, and values of the shape parameter are undefined in several cases. This feature has been already known, i.e. Azzalini and Arellano-Valle showed in [2] that the ML estimates of the shape parameter can be diverged for finite samples with non-negligible probability, especially in the cases when true value of the parameter is close to its boundaries. On the other hand, their suggested bias pre-correction cannot be implemented straightforwardly in our case.

In the present research we intend to improve the models in order to prevent bias of the estimates and divergence of the shape parameter.
Keywords: gravity model, ML estimates, bias, skew-normal distribution.
References:
[1] Andronov A., Santalova D. Statistical estimates for modified gravity model by aggregated data. Communications in Statistics - Simulation and Computation, Taylor & Francis, 2012, 41:6, pp. 730 – 745.
[2] Azzalini, A., Arellano-Valle, R.B.: Maximum Penalized Likelihood Estimation for Skew-Normal and Skew-t distributions. Journal of Statistical Planning and Inference, 2013, 143, pp. 419–433.
[3] Santalova D. Experimental Analysis of Statistical Properties of Parameter Estimation of Modified Gravity Model. Automatic Control and Computer Science, 2013, 47:2, pp. 99 – 106.
[4] Santalova D. Migration Flows Estimation Using the Modified Gravity Model. In: 15th ASMDA International Conference, 25-28 June 2013, Mataró, Spain. Christos H. Skiadas (Eds.). International Society for the Advancement of Science and Technology, 2013, pp. 186.

The contribution of 'avoidable' causes of death to life expectancy gains in Portugal

Paula Santana, Ângela Freitas, Ricardo Almendra, Rui Gama, Adriana Loureiro
Universidade de Coimbra, Portugal

Avoidable mortality is an appropriate indicator for monitoring health policies and services. It corresponds to causes of death selected from a cluster of deaths considered to be avoidable given the existing possibilities for primary prevention and medical care.
The aim of this study is: i) to assess the contribution of amenable causes of death due to primary prevention and health promotion, to medical care and ischaemic heart disease (partly primary prevention, partly medical care indicator) upon life expectancy in Portugal in the last 20 years, considering three periods (1989-1993, 1999-2003 and 2008-2012); ii) estimate and compare the potential gains in years of life through the partial or total elimination of these amenable deaths.
To assess the impact of "avoidable" causes of death on life expectancy, Multiple Decrement Life Tables were constructed, estimating the likelihood of death in each age group (<1, 1-4,5-9,…, 75 years and more). The cluster of causes of death considered avoidable was set up according to Nolte, et al (2004).
Although decreasing, the avoidable mortality is still considered an important burden to the Portuguese life table. Over the 20 years studied, the proportion of amenable deaths is lower in women, and the difference within sex is growing once the decreasing trend in women is steeper than in man.

Between 1989-93 and 2008-12, life expectancy at birth increased by 5.2 years in men and 5.9 years in women, being in 2008-12, 77.89 and 85.02, respectively. For instance, it was estimated that, if deaths amenable to primary prevention and health promotion decreased 50%, life expectancy at birth would increase 0.84 years in man and 0.31 in women. With a 100% decrease, in 2008-12, life expectancy at birth would be 79.55 for men and 85.63 for women. Even so, the potential gains in life expectancy would be higher if the amenable death due to medical care and to ischaemic heart disease were removed.
The causes of premature death affecting the population result from behaviours and attitudes towards health and the use of healthcare services that are determined (both positively and negatively) by the environment in which individuals live. There needs to be a reinforcement of public policies that encourage healthy practices, particularly as regards territorial organization.
Key Words: Avoidable mortality; Life expectancy; Primary prevention; Medical care

Error orthogonal models: building up complex models

Carla Santos[1], Célia Nunes[2], João Tiago Mexia[3]
[1]College of Technology and Management, Polytechnical Institute of Beja, Department of Mathematics, CMA- Center of Mathematics and its Applications (New University of Lisbon), [2]University of Beira Interior, Department of Mathematics and Center of Mathematics, [3]Faculty of Science and Technology, New University of Lisbon, Department of Mathematics and CMA- Center of Mathematics and its Applications, Portugal

Error orthogonal models, EO, are a particular case of models with orthogonal block structure, OBS, where the least squares estimators, LSE, for estimable vectors are uniformly best linear unbiased estimators, UBLUE.
Using the algebraic structure of the error orthogonal models, based on commutative Jordan algebras constituted by symmetrical matrices, CJAS, and the Cartesian product we build up complex models from simpler ones through crossing, nesting and joining.
When a single model is studied, we say that the factors of this model are crossed when every level of one factor occurs with every level of the other factor. So the treatments are formed as the combinations of all levels of the factors

Crossing u models, each one with only one factor with a_1, K, a_u levels, we obtain the same combination of levels we had crossing u factors of one model, thus the same number of treatments.
Analogously, nesting u factors of a model is equivalent to nesting u models each one with only one factor.

Joining mixed models we obtain models whose observations vector is the overlap of the observations vectors of the initial models.
We prove that when the initial models are error orthogonal with segregation or matching so is the model obtained through joining.
Keywords: Mixed models, COBS, EO, Error orthogonal models, Models crossing, Models nesting, Models joining, Commutative Jordan algebras.
References
1.Jesus, V., Fonseca, M., Mexia, J.T. & Zmyslony, R. (2009b). *Binary operations and canonical forms for factorial and related models*. Linear Algebra and its Applications, 43, 2781–2797.
2. Mexia, J. T., Vaquinhas, R., Fonseca, M., Zmyslony, R. (2010). *COBS: Segregation, Matching, Crossing and Nesting*. Latest Trends on Applied Mathematics, Simulation, Modeling, 4th International Conference on Applied Mathematics, Simulation, Modelling (ASM "10)
3.Santos, C., Nunes, C. & Mexia, J.T. (2007b). *OBS, COBS and Mixed Models associated to commutative Jordan Algebra*. In proceedings of 56th session of the International Statistical Institute. Lisboa.
4. Santos, C. (2012) *Error ortogonal models. Structure, operations and Inference*. PhD Thesis, UBI – Covilhã.
5. VanLeeuwen, D.M., Seely, J.F. & Birkes, D.S. (1998). *Sufficient conditions for orthogonal designs in mixed linear models.* Journal of Statistical Planning and Inference, 73, 373-389.
6. Zmyslony, R. (1978). *A characterization of best linear unbiased estimators in the general linear model.* Lecture Notes in Statistics, 2, 365-373.

Analytical Results For Nonlinear Finite Dimensional Feedback Particle Filters

Guillaume Sartoretti[1], Max-Olivier Hongler[1], Roger Filliger[2]
[1]Swiss Federal Institute of Technology in Lausanne, [2]Berner Fachhochschule, Switzerland

As recently shown in [1], a natural connection exists between nonlinear filtering (NF) using feedback particle filters (FPF) and stochastic differential game theory (in particular "Mean field Games"(MFG). Here we analytically explore this connection by considering the following scalar NF situation:

$$\begin{cases} dX(t) = f_\mu(X(t))dt + \sigma_s dW_s(t), & (nonlinear\ system\ driven\ by\ WGN) \\ dZ(t) = hX(t) + \sigma_o dW_o(t), & (linear\ system's observation\ driven\ by\ WGN) \end{cases} \quad (1)$$

where $dW_s(t)$ and $dW_s(t)$ are two independent White Gaussian Noise (WGN) processes, $\sigma_s\ (resp.\ \sigma_o)$ is the diffusion constant of the system noise (resp. observation noise), and $f_\mu(x)$ is a nonlinear drift

parametrized by a control parameter μ. When $f_\mu(x)$ is nonlinear, the transition probability density (TPD) solving the associated Fokker-Planck equation (FPE) is non Gaussian and the filtering operations to propagate involve, in general, an infinite hierarchy of moments (i.e. we have an infinite dimensional problem). To realize the filtering task, it is essentially required to actually solve the FPE, and to this aim FPF algorithms use a swarm of *N* particles $X_k(t), k = 1, 2, \cdots, N$ subject to an optimal control algorithm $U_k(t, X_k(t), \mathrm{X}(t))$ with dynamics:

$$\begin{cases} dX_k(t) = f_\mu(X(t))dt + \sigma_s dW_{k,s}(t) + dU_k(t, X_k(t), \mathrm{X}(t)), \\ \mathrm{X}(t) := (X_1(t), X_2(t), \cdots, X_N(t)), \end{cases} \tag{2}$$

where $U_k(t, X_k(t), \mathrm{X}(t))$ minimizes the Kullback-Leibler distance between the exact and the measured system's PDFs [1]. As exposed in [2,3], for a specific class of f_μ's satisfying a Riccati equation, the filtering problem remains finite dimensional, a property that has been exploited in [4]. As suggested by [4], we exhaustively calculate the explicit filters belonging to the Beneš' class and obtain compact forms for the TPD and its moments. We emphasize that, for a selected range of μ, the potential $V_\mu(x)$, from which the Beneš' drift derives $(f_\mu(x) := -\frac{d}{dx}V_\mu(x))$, exhibits a double barrier shape implying that the system in Eq.(1) has metastable dynamics [3]. Thanks to the exact non Gaussian TPD, we are able to analytically express the optimal control $U_k(t, X_k(t), \mathrm{X}(t))$ for both discrete and continuous time measurements. The exact solution can, in particular, be compared with approximations given by extended Kalman and other approximate algorithms devised to tackle nonlinear systems' evolutions. With [1], we observe that for $N \to \infty$, the filtering problem Eq.(2) can be viewed as a nonlinear (and hence non Gaussian) Mean-Field differential Game (MFG), for which we are here able to write the exact and very explicit transient solution.

References

1. T. Yang, P. G. Mehta, and S. P. Meyn. Feedback particle filter. IEEE Transactions on Automatic Control, 58(10):2465{2480, 2013.
2. Vaclav E. Beneš. Exact finite dimensional filters for certain diffusion with nonlinear drifts. Stochastics, 5:65{92, 1981.
3. Max-Olivier Hongler. Study of a class of nonlinear stochastic process - boomerang behavior of the mean. Physica D, 2:353{369, 1981.
4. Frederick E. Daum. Exact finite-dimensional nonlinear filters. IEEE Transactions on Automatic Control, AC-31(7):616{622, 1986.

A Finite Mixture Model with Trajectories Depending on Covariates

Jang SCHILTZ[1], Jean-Daniel GUIGOU[1], Bruno LOVAT[2]
[1]University of Luxembourg, Luxembourg School of Finance, Luxembourg, [2]University of Lorraine, BETA UMR CNRS, France

The current generalizations of Daniel Nagin's finite mixture model are not flexible enough for a lot of practical applications since they just allow an influence of the covariates on the intercept of the trajectories, enforcing thus a parallel behavior of the different subgroups defined by the covariates.
We present a new finite mixture model which overcomes these weaknesses. After passing in review the statistical test hypotheses that can be addressed by it, we show an application of the model to the analysis of the salary structure in Luxembourg.
Moreover, we show some features of our new R-package on finite mixture models.
Keywords: Finite mixture model, test hypotheses, R-package.

Modeling and analysis of cyclic inhomogeneous Markov processes: a wind turbine case study

Teresa Scholz[1], Frank Raischel[2], Pedro Lind[3], Vitor Lopes[1]
[1]National Laboratory for Energy and Geology (LNEG), Portugal, [2]Center for Geophysics, IDL, University of Lisbon, Portugal, [3]ForWind- Center for Wind Energy Research, Institute of Physics, Carl-von-Ossietzky University of Oldenburg, Germany

This paper presents the development of a cyclic inhomogeneous Markov model based on a multi-dimensional state space. Each time-dependent transition probability is expressed as a Bernstein polynomial. The model parameters are estimated by solving a constrained optimization problem: The objective function combines two maximum likelihood estimators, one to ensure that the Markov process long-term behavior reproduces the modeled data accurately and another to capture deviations from this average behaviour. A convex formulation for the overall optimization problem is presented and its applicability demonstrated through the analysis of a case-study: The methodology is applied to a three-year dataset (wind power, speed and direction) collected from a wind turbine located in a mountainous region in Portugal. It is shown that the model is capable of reproducing the daily patterns present in the three variables of the original data.

Statistical Estimation of Quadratic Density Functionals for Stationary *m*-Dependent Sequences

Oleg Seleznjev[1], David Källberg[1], Nikolaj Leonenko[2]
[1]Department of Mathematics and Mathematical Statistics, Umea University, Sweden, [2] School of Mathematics, Cardiff University College, UK

We study estimation of certain integral functionals for the marginal densities of stationary *m*-dependent sequences. One of the examples of such functionals is the Rényi entropy for the marginal distribution. The Rényi entropy is a generalization of the Shannon entropy and is widely used in mathematical statistics and applied sciences for quantifying the uncertainty in a probability distribution. We consider estimation of the quadratic Rényi entropy and related functionals for the marginal distribution of a stationary m-dependent sequence. The *U*-statistic estimators under study are based on the number of ε-close vector observations in the corresponding sample. A variety of asymptotic properties for these estimators are obtained (e.g., consistency, asymptotic normality, Poisson convergence). The results can be used in diverse statistical and computer science problems whenever the conventional independence assumption is too strong (e.g., ε-keys in time series databases, distribution identification problems for dependent samples).

Keywords: Entropy estimation, m-dependent sequence, U-statistics, quadratic Rényi entropy

References

1. Källberg, D., Seleznjev, O., Estimation of entropy-type integral functionals, Communications in Statistics - Theory and Methods (to be printed) (arXiv:1209.2544) (2012).
2. Källberg, D., Leonenko, N., Seleznjev, O., Statistical inference for R´enyi entropy functionals, Lecture Notes in Comput. Sci., 7260, 36-51 (2012).

Adaptive Stochastic Modeling Framework as Applied to Identification of a Simple Thermal Homeostasis Stochastic Model

Innokentiy V. Semushin, Julia V. Tsyganova, Elena S. Petrova, Anatoli G. Skovikov
Ulyanovsk State University, Russia

The applicability of our Active Principle of Adaptation (APA) to bioinformatics problems in general and to studying the thermal homeostatic mechanism in particular has been shown in our paper at The ASMDA 2013 Conference.
Warming up to the topic, we construct in detail the new general purpose APA-based parameter identification framework for discrete-time LQG

systems with control and filtering. We show how this framework can be applied to parameter identification of a simple human body daily temperature variation (HBDTV) stochastic model. The results of numerical experiments with the HBDTV data are included to confirm the validity of the proposed framework. Thereunder, a Convergence Indicating Assay Method (CIAM) for APA-based parameter identification processes is formulated.
This work is supported by the RFBR research grant No. 13-01-97035.
Key Words: Adaptation, active principle, homeostasis, parameter estimation, stochastic modeling, thermoregulation

Out of the crisis in Belarus and Russia: commonality and specificity of mortality changes

Victoria G. Semyonova, Tamara P. Sabgayda
Federal Research Institute for Health Organization and Informatics, Russia

We compare mortality dynamics in different age groups for period 1990-2011 as well as life expectancy in Russia and Belarus, where the socio-economic changes were carried out using the different models.
Key Words: Mortality dynamics, Age structure of mortality, Mortality structure, Life expectancy, Influence of socio-economic changes on mortality, Death causes

Near-critical Bienaymé-Galton-Watson processes escaping extinction

M. Conceição Serra[1], Serik Sagitov[2]
[1]Center of Mathematics, Minho University, Portugal, [2]Chalmers University of Technology, Sweden

In this talk we consider near-critical Bienaymé-Galton-Watson processes and propose a flexible way for building the skeletons.
Skeletons of branching processes are usually defined as trees of lineages and characterized by an appropriate feature that ensures future reproduction success. In the supercritical case a natural choice is to look for the lineages that survive forever. In the critical case it was earlier suggested to distinguish the particles with the total number of descendants exceeding a certain threshold. These two definitions lead to asymptotic representations of the skeletons as either pure birth process (in the slightly supercritical case) or critical birth-death processes (in the critical case conditioned on exceeding a high threshold value). The limit skeletons reveal typical survival scenarios for the underlying branching processes.

In this work we adress the near-critical case. Each vertex of the family tree is independently marked with a small probability. The branch connecting the root with a marked vertex is called a marked branch. The marked branches form a subtree of the family tree of the branching process and this will be called a skeleton. Such a skeleton is approximated, when marking is a rare event, by a birth-death process.
Keywords: Bienaymé-Galton-Watson process, decomposable multi-type process, birth and death process
References
1. O'Connel, N. "Yule process approximation for the skeleton of a branching process", J. Appl. Prob. 30, 725-729 (1993)
2. Sagitov, S. "Limit skeleton for critical Crump-Mode-Jagers branching processes", In Classical and modern branching processes, IMA Vol. Math. Appl., 84, pp 295-303, Springer, New York (1997)
3. Sagitov, S., Serra, M.C. "Bienaymé-Galton-Watson processes escaping extinction Adv. Appl. Prob. 41 225-246 (2009)

Estimating multi-factor discretely observed Vasicek term structure models with non-Gaussian innovations

Takayuki Shiohama
Faculty of Engineering, Tokyo University of Science, Japan

In this paper, we propose a multi-factor model in which the discretely observed short-term interest rates follow a non-Gaussian and dependent process. The state space formulation has the advantages of taking into account both the cross-sectional and time-series restrictions on the data and measurement errors in the observed yield curve. Clarifying the non-Gaussianity and dependency of the dynamics of short-term interest rates, we show that these features are important to capture the dynamics of the observed yield curve.
Key Words: Term structure, Vasicek model, state-space model

Probabilistic Approach for Comparing Partitions

Osvaldo Silva[1], H. Bacelar-Nicolau[2], Fernando C. Nicolau[3], Áurea Sousa[4]
[1]Dep. of Math., CES-UA, CMATI, University of Azores, [2]Faculty of Psychology, LEAD; ISAMB, CEA; University of Lisbon, [3]FCT, Department of Mathematics, New University of Lisbon, [4]Dep. of Math., CEEAplA, CMATI, University of Azores, Portugal

The comparison of two partitions in Cluster Analysis can be performed using various classical coefficients (or indexes) in the context of three

approaches (based, respectively, on the count of pairs, on the pairing of the classes and on the variation of information). However, different indexes usually highlight different peculiarities of the partitions to compare. Moreover, these coefficients may have different variation ranges or they do not vary in the predicted interval, but rather only in one of their subintervals. Furthermore, there is a great diversity of validation techniques capable of assisting in the choice of the best partitioning of the elements to be classified, but in general each one tends to favor a certain kind of algorithm. Thus, it is useful to find ways to compare the results obtained using different approaches. In order to assist this assessment, a probabilistic approach to comparing partitions is presented and exemplified. This approach, based on the VL (Validity of Linkage) Similarity, has the advantage, among others, of standardizing the measurement scales in a unique probabilistic scale. In this work, the partitions obtained from the agglomerative hierarchical cluster analysis of a dataset in the field of teaching are evaluated using classical and probabilistic (of VL type) indexes, and the obtained results are compared.

Keywords: Hierarchical cluster analysis, comparing partitions, affinity coefficient, VL methodology

Spectral analysis of Markov chains and graphs for ranking of evolving data and information

Sergei Silvestrov
Mälardalen University, Sweden

Spectral analysis of nonnegative matrices play fundamental role for the theory of Markov chains and the spectral graph theory. The common foundation is the Perron-Frobenius theory for matrices with non-negative entries. In this article we will present some of the main results and notions of this fruitful interplay in context of ranking and relevance analysis in complex evolving data and information structures as well as some novel approaches to ranking based on spectral properties of matrices and Markov chains, in particular modifying and extending the famous PageRank algorithm. We will also present new results and a number of new open problems in this direction.

Keywords: Perron-Frobenius theory, Markov chain, spectrum, matrix, graph, PageRank

How to find the State of a System by Inverting the Hitting Time Probability Density Function

[1]Christos H Skiadas, [2]Charilaos Skiadas
1ManLab, Technical University of Crete, Greece, 2Department of Mathematics and Computer Science, Hanover College, USA

Consider the stochastic process S_t with drift μ_t (W_t is the Wiener process)

$$dS_t = d\mu_t + dW_t \tag{1}$$

which we immediately integrate to give the stochastic paths S_t

$$S_t = \int_0^t \mu_s ds + \int_0^t dW_s = H_t + \int_0^t dW_s \tag{2}$$

where H_t is referred to as the State Function of the system.
For a linear form of H_t the probability density function $g(t)$ for the first exit time or hitting time is of the form (k is a constant):

$$g(t) = \frac{k}{\sqrt{2\pi\, t^3}} e^{-\frac{(H_t)^2}{2t}} . \tag{3}$$

The aim is to find the unknown function H_t given the distribution function $g(t)$ from the inverse formula:

$$H_t = \pm \left| \left(-2t \ln \frac{g(t)\sqrt{2\pi t^3}}{k} \right)^{1/2} \right| . \tag{4}$$

Schrödinger and Smoluchowsky have introduced the so-called Inverse-Gaussian (3) in the same Journal Issue in 1915. This is the simplest case for a system with a linear state function $H_t=l-bt$ (b, l are parameters). When the barrier is set to zero the parameter $k=l$ and a very simple case arises.

For a continuous declining function which crosses zero it was proven (Skiadas and Skiadas, 2013a, 2013b, 2014) that the parameter $k=k_{max}$, where k_{max} is the maximum value of $k(t)$ provided from the form

$$k_{est} = \max\left(\sqrt{2\pi t^3}\, g(t)\right) \tag{5}$$

thus making possible the estimation of H_t via (4).

However, the formula (3) is an approximation of the model proposed by Jennen and Lerche (1981) for a nonlinear function H_t

$$g(t) = \frac{\left| H_t - tH_t' \right|}{\sqrt{2\pi t^3}} e^{-\frac{(H_t)^2}{2t}} \tag{6}$$

It is clear that (6) cannot be solved for the unknown state function $H(t)$ given $g(t)$. However, by adding a correction term f_t we can find an approximation of (6) of the form

$$g(t) = \frac{k}{\sqrt{2\pi t^3}} e^{-\frac{(f_t + H_t)^2}{2t}} \tag{7}$$

Now the inversion of (7) will provide immediately the following form (k_{est} is found from (5))

$$H_t + f_t = \pm \left| 2t \ln\left(\frac{\sqrt{2\pi t^3}\, g(t)}{k_{est}} \right) \right| \tag{8}$$

When data sets for $g(t)$ are provided for a convenient time period, then (H_t+f_t) is estimated (f_t is found by stochastic simulations, Skiadas and Skiadas 2014).

In the latter case it is easy to find the State of a System as the Health State of a Population from the probability density of the death process at age x (Janssen and Skiadas, 1995, Skiadas and Skiadas, 2013a, 2013b, 2014) expressed by D_x or other demographic and actuarial characteristics as the mortality expressed by m_x or q_x. The Health State presented here is analogous with the Health Status of a Population a terminology used by Chiang (1966), Sullivan (1966, 1971) and Torrance (1976). A study including all the World Health Organization countries is presented. The implications to Demographic and Actuarial Science are straightforward.

Keywords: Stochastic modeling, First exit time probability density function, Hitting time model, Health State Function, Health Status of Population, Inverse Gaussian, Life Table Data, Inversion.

References

C. L. Chiang, An Index of Health: Mathematic Models. *PHS Publication* No. 1000, Series 2, No. 5 (1966).

J. Janssen and C. H. Skiadas, Dynamic modelling of life-table data, Applied Stochastic Models and Data Analysis, 11, 1, 35-49 (1995).

C. Jennen and H.R. Lerche, First exit densities of Brownian motion through one-sided moving boundaries. *Z. Wahrsch. uerw. Gebiete*, 55: 133-148 (1981).

C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*, ISAST, Athens, (2nd Ed., January 2013) http://www.amazon.com/Health-State-Function-

Population/dp/6188046505/ref=sr_1_1?s=books&ie=UTF8&qid=1364343495&sr=1-1
C. H. Skiadas and C. Skiadas, *Supplement: The Health State Function of a Population*, ISAST, Athens, (December 2013). http://www.amazon.com/Supplement-Health-State-Function-Population/dp/6188069831/ref=sr_1_1?s=books&ie=UTF8&qid=1391727448&sr=1-1&keywords=Supplement%3A+The+Health+State+Function+of+a+Population
C. H. Skiadas and C. Skiadas, The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods,* 43: 1585-1600 (2014).
D. F. Sullivan, Conceptual Problems in Developing an Index of Health, PHS Publication No. 1000, Series 2, No. 17 (1966).
D. F. Sullivan, A single index of mortality and morbidity. *HSMHA Health Reports*, 86:347-354 (1971).
G. W. Torrance, Health Status Index Models: A Unified Mathematical View, *Management Science*, 22, 9 (1976).

A Further Analysis on Exploring Sullivan's Health Status Index of Mortality and Morbidity

Christos H Skiadas[1], Charilaos Skiadas[2]
[1]ManLab, Technical University of Crete, Greece, [2]Department of Mathematics and Computer Science, Hanover College, USA

In a previous paper we have explored the Sullivan's health status index and we have proposed a method to estimate the Active Life Expectancy (ALE) without data from questionnaires but with only death and population data provided by the bureau of the census. In this paper we expand our method in order to estimate the Healthy Life Expectancy for all cases of disability (severe, moderate and light).
The model used is of the following form:

$$g(t) = k(l + (c-1)(bt)^c)(t)^{-3/2} e^{-\frac{(l-(bt)^c)^2}{2t}}$$

where $g(t)$ is the hitting time probability density function at time t, and b, l, c, k are parameters. The Health State Function is

$$H(t)=l-(bt)^c$$

and is presented in the following Figure 1. The declining curve is the estimated from the hitting time procedure health state function $H(t)$ of the

population. The line AB expresses the optimum case with no disabilities and other health declining phenomena whereas the curve ADC expresses the real situation.

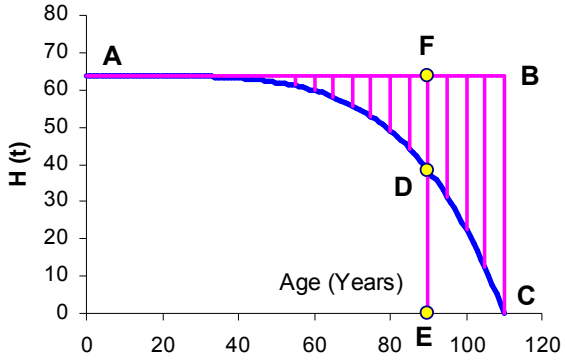


Figure 1. Health State of Population

That is immediately clear is that the area ABCDA represents the total loss of health of the population whereas the area ADCOA expresses the Healthy region. As a consequence the fraction ABCDA/ABCOA provides the net loss of health of the population from zero to 110 years of age. This is a very important information enabling us to estimate the Active Life Expectancy (ALE) and the Loss of Active Life Years (LALY) according to the terminology used by the Ministry of Health and Statistics New Zealand (2008) in a published discussion paper.
The estimates at an age level $x=t$ provide the formula for the fraction of the loss of active life years px

$$px=(DF)/(EF)$$

It is clear that $px=0$ at zero age and 1 at the end of the life time. The resulting fraction of the loss of active life years during age is presented in Figure 3. For our application we use the classical Life Table expanded for the estimation of the Active Life Expectancy. The method of Sullivan is applied.

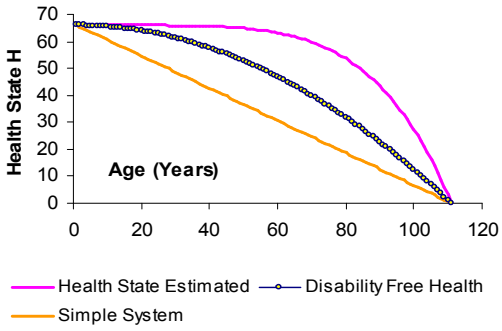


Figure 2. Health State for various cases

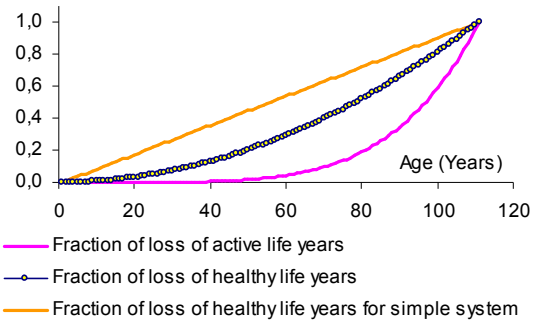


Figure 3. Fractions of loss of healthy life years for various cases

We further expand Figure 1 to include the Health State for a "Simple System" which is expressed by a straight declining line and a disability free health case illustrate with a small circle curve within the simple system and the estimated (see Figure 2). The intermediate curve H_{est} is estimated from the estimated curve H_{est} for the health state by using the following formula:

$$H_{int}=H_{est}-\alpha t^{\beta}$$

where α and β are parameters. Only the parameter β is needed as the parameter α is estimated so that H_{int} (t=110)=0. β=0 characterizes the "simple case" expressed by the straight line whereas the case without disability β tends to infinity (see the line AFB in Figure 1).
The resulting fractions of the loss of healthy life years for the three cases are illustrated in Figure3. Based on this method we estimate the healthy life expectancy for UK from 2005-2011 and compare it with the Eurostat estimates.

TABLE I

Estimated Healthy Life Years for UK

Males	2005	2006	2007	2008	2009	2010	2011
UK Leb	76.9	77.1	77.3	77.5	78.0	78.3	78.7
UK Eurostat	64.2	64.8	64.6	65.0	65.0	65.0	65.2
UK estimated	63.1	63.2	63.4	63.5	63.8	64.0	64.3
UK simple	49.3	49.4	49.4	49.5	49.7	49.8	49.9
UKmax loss	27.5	27.7	27.9	28.0	28.3	28.5	28.8

Keywords: Stochastic modeling, First exit time probability density function, Hitting time model, Health State Function, Health Status of Population, Life Table Data, Sullivan.

References

C. L. Chiang, An Index of Health: Mathematic Models. *PHS Publication* No. 1000, Series 2, No. 5 (1966).

Ministry of Health and Statistics New Zealand. (2008). Health Expectancy: Toward Tier 1 official statistic status. Wellington: Ministry of Health and Statistics New Zealand. *Discussion paper*, November 2008.
C. H. Skiadas and C. Skiadas, Development, Simulation and Application of First Exit Time Densities to Life Table Data. *Communications in Statistics* 39, 444-451 (2010)
C. H. Skiadas and C. Skiadas, *The Health State Function of a Population*, ISAST, Athens, (2nd Ed., January 2013) http://www.amazon.com/Health-State-Function-Population/dp/6188046505/ref=sr_1_1?s=books&ie=UTF8&qid=1364343495&sr=1-1
C. H. Skiadas and C. Skiadas, *Supplement: The Health State Function of a Population*, ISAST, Athens, (December 2013). http://www.amazon.com/Supplement-Health-State-Function-Population/dp/6188069831/ref=sr_1_1?s=books&ie=UTF8&qid=1391727448&sr=1-1&keywords=Supplement%3A+The+Health+State+Function+of+a+Population
C. H. Skiadas and C. Skiadas, Exploring Sullivan's Health Status Index of Mortality and Morbidity, *Chaotic Modeling and Simulation*, 4: 511-518 (2013).
C. H. Skiadas and C. Skiadas, The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods,* 43: 1585-1600 (2014).
D. F. Sullivan, Conceptual Problems in Developing an Index of Health, PHS Publication No. 1000, Series 2, No. 17 (1966).
D. F. Sullivan, A single index of mortality and morbidity. *HSMHA Health Reports*, 86:347-354 (1971).
G. W. Torrance, Health Status Index Models: A Unified Mathematical View, *Management Science*, 22, 9 (1976).

Unemployment in pre-retirement ages and disease onset in Europe and in the U.S.

Aïda Solé-Auró, Emmanuelle Cambois
Ined – Institut National d'Études Démographiques, France

Introduction: The largest mortality gap between Europe and the U.S. appears to be at the ages immediately preceding retirement and preliminary evidence suggests that there is no widening of the gap at older ages (Crimmins et al. 2011). Because mortality results from an accumulation of processes over a lifetime, the US-Europe gap in these ages must be explained by differences in both early life conditions and health dynamics during the working years. Several factors may explain international health differences at the country levels including general exposure to adverse conditions, living and work conditions, level of medical care and access to care, support for unemployment and family social security, health practices, or education system (Mackenbach et al. 2008). Conditions of employment and unemployment during the working years may be an important factor in health differences in pre-retirement period. Unemployment is associated with increased morbidity. Poor health might be a reason for job loss, but unemployment can also induce health deterioration. Job loss can translate into a reduction of income, material resources and access to care in the same time than a reduction in the well-being and social network with possible increases of the risk of psychological distress. But all these losses related to unemployment

vary across countries due to the social protection system. Indeed, there were evidences that differences in welfare systems could explain country variations in the association between self-reported health and unemployment, at least for women (Bambra and Eikemo 2009). In this paper, we analyze the impact in the social protection system on the association between unemployment and health by looking not only at the prevalence but also at the onset of diseases in the European and the U.S. contexts over time to see whether health gaps in the pre-retirement period might be linked to different systems.

Data: This analysis uses the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Health and Retirement Study (HRS) to compute prevalence and incidence of health conditions at individual level using panel data (from 2004 to 2010). This study includes information about individuals aged 50 to 80 for eleven European countries and the U.S.

Methodology: The use of panel data from SHARE and HRS is to estimate health trajectories after age 50. We focus on the most common and frequent health conditions, which are major causes of mortality at older ages: heart disease, stroke, lung disease, diabetes, hypertension, and cancer. We first examine descriptive data on prevalence and incidence of diseases by age groups in Europe and the U.S. Then we examine 6-year incidence rates among those who did not have a condition at the initial interview. We estimate incidence among those individuals that are unemployed – or become unemployed during the period - using hazard models for each of the six diseases and we include in a first step, socio-demographic factors, and in the second step, macro-indicators, to determine how these factors relate to disease onset and how differences affect the differences within Europe and between Europe and the U.S.

Results and Conclusions: Eurostat data indicate variability in the unemployment rate across Europe and over time. Except for Spain, the unemployment rates in 2010 range from 4% to 12% within Europe. Lengthy unemployment of 1 year or longer has been increasing in Spain, the U.K and U.S. Unemployment and disability rates are related. Individuals with long periods of unemployment may exhaust unemployment benefits, and may transfer to disability benefits. The prevalences of disease are significantly higher in the U.S. than in Europe for each of these conditions. Close to one fifth (18.3%) of Americans have heart disease in this age range compared to one-tenth (10.4%) of the Europeans. Prevalence of most conditions is between 40% and 100% higher in the U.S. compared to Europe. A review of age specific incidence shows that the incidence for many conditions post age 50 is worse in European countries. Our next step is to further analyze the interaction with unemployment using individual data over time (health transitions) and by countries. Moreover, there will be a selection of relevant macro-level factors, and structural and sustainable indicators, for every country to cover broad areas of labor-force participation (long-term unemployment rate, mid-adult ages unemployment rate), wealth and care expenditure, which we can obtained from the Eurostat website.

Discrepancies of Life Tables

Eugeny Soroko
Institute of Demography, National State University Higher School of Economics

Several types of discrepancies in human life tables are put under consideration in this paper. Compared with well-known problems caused by inaccuracy of vital statistics or appeared in small populations, a set of new ones is analyzed. They are attributable to rounding of death rate decimal digits at the stage of publication, random fluctuations of death counts in case of very small probabilities. Special attention is paid to zero qx and phenomena suggested to name “pseudo-discreteness”, “pseudo-equality”, “pseudo-growth” of indicators in life tables for population in developed countries.
Keywords: Life table, Death rate, Population, Rounding, Probability of death

Analyzing elastic body wave conversion data

D. A. Sotiropoulos
Technical University of Crete, Chania 73100, Greece

It is known that several materials such as cork with Portugal being the largest world producer of it, foam, glass, concrete and auxetic materials having a Poisson’s ratio in the range [-1, 0.263] allow complete conversion of elastic body waves. That is, upon incidence on a planar surface at a specific angle which depends on the Poisson’s ratio of the material, shear waves reflect only as longitudinal waves and longitudinal waves reflect only as shear waves. Based on this property, non-destructive ultrasonic evaluation of these materials finds industrial applications during or after their production. In such evaluations, the structural integrity of the material is evaluated from the relationship between Poisson’s ratio and angle of incidence for complete body wave conversion. When there are is no damage in the material, the quotient (1-2v)/[2(1-v)] where v is the Poisson’s ratio of the material, may be expressed in closed form as the quotient of a third degree polynomial in m to a first degree polynomial in m, where m= (1/sinθ)2 and θ is the angle of incidence for complete conversion of a shear wave. This relationship shows that determining m from v whose exact value may not be known due to possible structural deficiencies in the material, is not straight forward. In the present paper an approximation is proposed for the relationship between v and m in order to facilitate material evaluation from complete body wave conversion ultrasonic data. The approximation is based on an analytical derivation which yields the exact relationship between v and m as m(1-2v)/[2(1-v)]=1+(m-2)4/[16(m-1)]. Cork-like materials have a diminishing Poisson’s ratio and, thus, m=2. The proposed approximation takes advantage of the slow departure of the product m(1-2v)/[2(1-v)] from m=2, therefore, proposing it to be equal to 1. The error involved in the approximation for structurally sound materials is (m-2)4/[16(m-1)], thus being equal to zero for cork-like materials. The proposed approximation results in errors that are less than 1% for materials with Poisson’s ratio in the range [-0.6, 0.2] for

which the exact m ranges from 1.47 to 2.69. The approximation's errors become considerably larger for v=(0.2, 0.263] or, equivalently, for m=(2.69, 3.56] where they vary from 0.8% to 15%. For auxetic materials with extreme negative values the errors of the approximation are much smaller increasing from 1% to 2.7% as Poisson's ratio v decreases from -0.6 and -1, or, equivalently, as m decreases from 1.47 to 1.37. In conclusion, an approximation for the relationship between Poisson's ratio and angle of incidence for complete elastic body wave conversion has been proposed which facilitates evaluation of the structural integrity of materials from ultrasonic data.

Hierarchical Cluster Analysis of Groups of Individuals: Application to Business Data

Áurea Sousa[1], Helena Bacelar-Nicolau[2], Osvaldo Silva[3]
[1]Dep. of Math., CEEAplA, CMATI, University of Azores, [2]Faculty of Psychology, LEAD; ISAMB, CEA; University of Lisbon, [3]Dep. of Math., CES-UA, CMATI, University of Azores, Portugal

In this work, classical as well as probabilistic hierarchical clustering models are used to look for typologies of *H* independent groups of individuals in two contexts: classical three-way data and symbolic/complex data. We present one example on the business area, in which the data are issued from a questionnaire in order to evaluate the satisfaction and quality with the services provided to customers by an automobile company. The Agglomerative Hierarchical Cluster Analysis is based on appropriated extensions of the basic affinity coefficient. The used probabilistic aggregation criteria - in the scope of the *VL* methodology (*V for Validity*, *L for Linkage*) - resort essentially to probabilistic notions for the definition of the comparative functions.
The clustering of independent groups of individuals was made according to two strategies: one based on a particular case of the generalized weighted affinity coefficient for the case of a classical three-way data table, and the other one based on the weighted generalized affinity coefficient for modal variables.
In the first approach, the data were initially represented in *H* sub-tables (one sub-table for each of the groups of individuals), containing, respectively, N_1, N_2,…, N_H, individuals described by *p* identical variables. Later *H* new sub-tables, each one containing the same number m= $min\{N_1, N_2, \ldots, N_h\}$ of individuals (selected from a stratified random sampling) had to be obtained from the initial corresponding sub-tables. Thus, in this case, we loss information because we can´t work with all the sample but only with a stratified random subsample. Contrary, in the second approach it is possible to work with the entire dataset.
The clustering results provided by both strategies were compared, and the differences found between the satisfaction typologies were due to the smaller number of individuals of each group when we apply the first

approach as a consequence of the sampling process. Nevertheless, we might have opted by inquiring a largest number of individuals in each group, during the planning of the investigation. We used the global statistics of levels (*STAT)* to evaluate the obtained partitions in the agglomerative hierarchical clustering.
Keywords: Hierarchical cluster analysis, Affinity coefficient, independent groups of individuals*, VL* Methodology, Three-way data, Symbolic data.

Applications of the Bayesian and the Trimmed Likelihood Estimation in Multitype Branching Processes

Ana Staneva
Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics, Department of Probability, Operations Research and Statistics, Technical University of Sofia, Bulgaria

In our work we consider illustrative examples of multitype branching processes, allowing for Bayesian estimation and the explicit calculation of the breakdown point of the trimmed likelihood estimators of the individual distributions of the process. The estimators use several independent family trees of the processes and their advantage in the presence of oultiers is shown via simulations and computational results.
Keywords: Multitype branching processes, Bayesian estimation, Trimmed likelihood, Breakdown point
* This work is supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014) and is partially supported by the financial funds allocated to the Sofia University “St. Kl. Ohridski”, grant No 012/2014.

Inference for Quantile Measures of Peakedness

Robert G. Staudte
Department of Mathematics and Statistics, La Trobe University, Australia

The peakedness measure of Horn (37:55-56, 1983, *The American Statistician*) is here extended to arbitrary densities so that it can detect bimodality. Ruppert (41:1-5.1987, *The American Statistician*) introduced a class of measures of peakedness defined as the ratio of two interfractile ranges pi_{q,r}=R_q/R_r, where R_r=x_{1-r}-x-r$ is the difference of two quantiles, and 0<q<r<0.5. He showed that pi_{q,r} satisfied certain convexity requirements expected of a measure of kurtosis and shared by the classical measure of kurtosis defined in terms of moments; however, the new measures were not affected by extreme outliers. Ruppert derived and compared the influence functions of these

and other measures of peakedness and kurtosis. He also showed that his measure was approximately monotone in Horn's measure for r near 0.5. This result continues to hold for the extension of Horn's measure. We describe distribution-free confidence intervals for pi_{q,r}. The methodology is similar to that utilized by the author in recent studies on the standardized median x_{0.5}/R_r (*STAT*, 2:184-196) and quantile skewness measures introduced by Bowley (1920, *Elements of Statistics*) and Hinkley (1975, *Biometrika*, 62:101-111). These results are largely distribution-free, and a formula for the sample size required to obtain accurate coverage of distribution-free confidence intervals for pi_{q,r} with a prespecified relative width are provided.

Optimal design for parameters of stochastic processes

Milan Stehlik
Departmento de Matematica, Universidad Tecnica Federico Santa Maria, Chile

The determination of optimal designs for models with a correlated errors is substantially more difficult and for this reason not so well developed. Stochastic process with parametrized mean and covariance is observed over a compact set. The information obtained from observations is measured through the information functional (defined on the Fisher information matrix). We will focus on efficient designs for the parameters of correlated processes. We will discuss the role of equidistant designs for the correlated process. Such designs have been proved to be optimal for parameter of trend of stationary Ornstein-Uhlenbeck process (see [2]). For such a process we provide a study of small sample and asymptotical comparisons of the efficiencies of equidistant designs with taking into account both the parameters of trend as well as the parameters of covariance. If only trend parameters are of interest, the designs covering more-less uniformly the whole design space are rather efficient when correlation decreases exponentially (see [2]). Some issues on designing for spatial processes (see [1,3,5]) will be also provided. Finally we will concentrate on relaxing the continuity of covariance (see [6]). We will introduce the regularity conditions for isotropic processes with semicontinuous covariance such that increasing domain asymptotics is still feasible, however more flexible behavior may occur here. In particular, the role of the nugget effect will be illustrated. During the talk we introduce an compound criterion (see [4]) and integrated compound criterion to discuss their potential of optimal designing for parameters of correlated processes. The case of planar OU sheet will be discussed (see [7]) and role of heteroscedasticity presented (see [8]). Finally, application in troposphere methane modelling will be illustrating the developed methods (see [9]).

Key Words: Optimal design, correlated processes, Ornstein-Uhlenbeck Process and Sheet

REFERENCES

[1] Dette H., Kunert J. and Pepelyshev A. (2008). Exact optimal designs for weighted least squares analysis with correlated errors, Statistica Sinica 18:135-154.
[2] Kiselak J. and Stehlik M. (2008), Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process, Statistics and Probability Letters 78, 1388-1396.
[3] Mueller W.G. and Stehlik M. (2009). Issues in the Optimal Design of Computer Simulation Experiments, Applied Stochastic Models in Business and Industry, 25:163-177.
[4] Muller W.G. and Stehlik M. (2010). Compound optimal spatial designs, Environmetrics, 21: 354-364.
[5] Stehlik M., Rodriguez-Diaz J.M., Mueller W.G. and L'{o}pez-Fidalgo J. (2008). Optimal allocation of bioassays in the case of parametrized covariance functions: an application in lung's retention of radioactive particles, Test 17,56-68.
[6] Stehlik M. (2009) Topological regularities for regression with correlated errors, Research Report.
[7] Baran, S., Sikolya, K. and Stehlik M. (2013), "On the optimal designs for prediction of Ornstein-Uhlenbeck sheets" Statistics and Probability Letters, Vol. 83 (6), 1580-1587.
[8] Boukouvalas A., Cornford D., Stehlik M. (2014), "Optimal Design for Correlated processes with input-dependent noise", Computational Statistics and Data Analysis, 71, 1088-1102.
[9] Rodriguez-Diaz J. M., Santos-Martin T., Waldl H. and Stehlik M. (2012), "Filling and D-optimal designs for the correlated Generalized Exponential models", Chemometrics and Intelligent Laboratory Systems, 114: 10-18.

Estimators in Discrete Time Multitype Branching Processes

Vessela Stoimenova
Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics, Department of Probability, Operations Research and Statistics, and Bulgarian Academy of Sciences, Institute of Mathematics and Informatics, Department of Probability and Statistics, Bulgaria

Branching processes form an important class of stochastic processes with numerous applications in different scientific and practical areas, many of them involving multitype modeling. Statistical estimation of the process' characteristics is an important issue in their study. In our work we review some commonly used estimators for multitype branching processes and study their behavior in the presence of outliers. One possible approach for estimation in a contaminated sample is to combine the asymptotic distributions of the estimators with the method for constructing robust estimators, based on the trimmed likelihood. We

consider numerical results for simulated data samples of different types - samples over family trees and over generation sizes - and compare estimators and their robust versions.
Keywords: Multitype branching processes, Estimators, Trimmed likelihood

* This work is supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014) and is partially supported by the financial funds allocated to the Sofia University “St. Kl. Ohridski”, grant No 012/2014.

Smoothing of probabilities of death for older people in life expectancy tables

Gustaf Strandell, Tomas Johansson
Statistics Sweden, Sweden

In 2012, Statistics Sweden conducted a review of the calculations for the Swedish life tables. The project had two main objectives:
1. A quality assured production system, and
2. Review of the handling of mortality risks for very old persons at the national and regional levels.
This paper focuses on the changes that were made in the handling of mortality risks for persons in the highest ages, where the population is small and mortality risks are high. It also shows how the changed method has affected the estimated average life expectancies at the national and regional levels.
Over the years, Statistics Sweden has used various methods to handle the weak relationship between mortality risks and age at the highest ages. Up to 1986, Wittstein's equalisation was used. which overestimated the mortality risks for the oldest persons. Therefore, Statistics Sweden created a method which basically used a "universal" equalisation curve that adjusted the level at the age of 91; thereafter an equalisation took place. Over time, did that method result in a systematic underestimation of mortality risks and thus an overestimation of the life expectancy for newborns, e0.
The revised equalisation method involves a procedure in SAS that adjusts the equalising function to the observed mortality risks. The procedure uses a least-squares method to adjust the specified function to the specified data material. By also taking into account the number of deaths by sex and age, the procedure assigns varying significance for the different age groups for the adjustment.
Mortality risks for the oldest persons were previously based on the national mortality risk regardless of the regional level. The disadvantage

of this approach was that regional differences for people in the highest ages were thus ignored. Nowadays the regional life tables are based on observations at the regional level.
Key Words: Lifetable equalisation mortality risks life expectancy

An Information Theoretic Measure of Cluster Validity for Fuzzy Clustering

Abdul Suleman
ISCTE - IUL Instituto Universitário de Lisboa, BRU-UNIDE, Portugal

We propose a new cluster validity index for fuzzy c-means clustering algorithm. It is based on the information theoretic principles with the same aim and spirit of AIC – Akaike's Information Criterion. We start to show the relationship between fuzzy partitions and unit simplices by means of a geometrical argument, and then convert it to an algebraic form. As a result, the observed data points are described by a convex linear combination of cluster prototypes. Looking at fuzzy partitions as a virtual coordinate system made it possible to design our index from high-level AIC-like functionals. This means we started from the loss or criterion function used to select predictors in a multiple regression analysis.
The way we measure the model bias is similar to the one in AIC; however, we realised that the measure of complexity used in this criterion, i.e. twice the number of parameters, does not fit the context of fuzzy c-means clustering. We alternatively balance the number of parameters by a measure of efficiency of fuzzy partitions, comparing the covariance structure of the observations to that induced by the decomposed data.
We tested the efficiency of the proposed index with synthetic and real data, specifically 100 independent samples of Normal-4 data and 15 datasets from UCI Machine Learning Repository. We compared the results with those provided by some popular available indices, namely the Xie and Beni index, the partition coefficient and exponential separation index, the modified partition coefficient index, and the cluster validation index based on relative degree of sharing. We did not see any substantial difference between all indices in case of Normal-4 data. The same cannot be said for more realistic data, where our index outperformed any of the others referred above. The empirical evidence positions it as a valid alternative to cluster validation problem in the context of fuzzy c-means clustering.
Key Words: Cluster validity; fuzzy c-means; information theory.

Expected, predictable and observed magnitudes of the blood pressures through 16 years after the man's stroke

Marijan Sviben
Freelance consultant, Republic of Croatia

It was May 11th 1997, when the man, being 65 years of age, suffering from hypertension, passed through the stroke. In the very beginning of recovery it was seen that magnitudes of systolic blood pressure (SBP) and of diastolic blood pressure (DBP) through the period longer than 15 days varied and their distributions could be presented according to the Pascal's triangle. Since it had been experienced much earlier that really reliable results of statistical elaboration could be obtained with 200 animal data at least but with 200-400 variables, the blood pressure measurement were taken at the same human being during the observation period from April 1st 1998 till May 8th 1999. The SBP magnitudes averaged 148.2 mm Hg and the DBP magnitudes 97.25 mm Hg. The mean of SBP measurements equal to 147.5 mm Hg and the average of DBP magnitudes as great as 97.55 mm Hg for males 60-69 years of age were reported after the research done decades ago. At one time the regression of systolic blood pressure on the age was presented with the coefficient of +1.38 mm Hg what meant that average magnitudes of SBP were greater year by year from 35 till 75 age of human beings. The expectations of the increase of the means of blood pressures in years following 67 in 1999 were established and proven from 2000 till 2002. In 2003 150.5 mm Hg for SBP was expected according to the equation of linear regression $Y_C = 148.78 + 0.71X_C$ (origin between 2001 and 2002, $X_C = 1$) but it was found 135.3 mm Hg. The SBP means were 128.8 mm Hg in 2004, 124.5 mm Hg in 2005 and 121.4 mm Hg in 2006. In 2003 101.5 mm Hg for DBP was expected according to the equation $Y_C = 98.588 + 1.153X_C$ but it was found 90.3 mm Hg. The DBP averages were 84.7 mm Hg in 2004, 84.2 mm Hg in 2005 and 75.6 mm Hg in 2006. Predictable values were derived from three equations for SBP and three for DBP magnitudes. The pharmaceutical companies were asked for the magnitudes of SBP and DBP after long period of the use of their remedies to be compared to predictable ones in 2011 for SBP 132.4, 81.8 or 177.4 mm Hg and for DBP 23.17, -554.3 or 74.94 mm Hg . The companies did not know the long term consequences of the use of their remedies. With the data observed till 2013 it was found the equation $Y_C = 126.44 - 4.60\ X_C + 0.85\ X_C^2$ (origin between 2004 and 2005, i.e. 72 and 73 years of age) resulted with the SBP magnitudes similar to observed ones till 2010. In 2013 predictable average of 82.08 mm Hg and observed one of 81.20 mm Hg for DBP were established.

Covariance and Correlation Swaps for Markov and Semi-Markov Volatilities

Anatoliy Swishchuk
University of Calgary, Canada

In this talk, we show how to price covariance and correlation swaps for financial markets with Markov-modulated volatilities. As an example, we consider stochastic volatility driven by two-state continuous Markov chain. In this case, numerical example is presented for VIX and VXN volatility indeces (S&P 500 and NASDAQ-100, respectively, since January 2004 to June 2012). Extension for semi-Markov volatilities is considered as well. (This is a joint talk with Giovanni Salvi, University of Rome 'La Sapienza', Rome, Italy)
Key Words: Markov volatility; semi-Markov volatility; covariance swap; correlation swap;

Lower bounds on the convergence rate of the Markov symmetric random search

Alexey Tikhomirov
Novgorod State University, Russia

The convergence rate of the Markov random search algorithms designed for finding the extremizer of a function is investigated. It is shown that, for a wide class of random search methods that possess a natural symmetry property, the number of evaluations of the objective function needed to find the extremizer accurate to ε cannot grow slower than |ln ε|.
Keywords: Random search, Global optimization, Stochastic optimization

Multichannel Queuing Systems with Impatient Customers and Regenerative Input Flow

Andrey Tkachenko
National Research University Higher School of Economics, Russia

In the present investigation we focus on the multichannel queuing system with r heterogeneous servers, common queue, regenerative input flow and impatient customers. Service times of customers are generally distributed. If a new arrived customer encountering j other customers in the system, then he joins the queue with probability f_j and balk with probability 1-f_j. We suppose that the sequence f_j is not increasing and f_j –> f >= 0. The adjoint customer stays in the line until his random waiting time w is over. In the present investigation we

assume that the random variable w is generally distributed and it can be improper one, i.e. a:=P(w=∞)>=0.
For this model we are planning to
1) establish necessary and sufficient condition for stability of the system,
2) prove functional limit theorem of convergence of the normalized queue length process to the diffusion process (reflected Brownian motion) under critically loaded conditions,
3) prove functional limit theorem of convergence of the normalized queue length process to the Brownian motion under overloaded conditions.
The main idea of proofs of these theorems is to introduce some auxiliary systems such that the queue-length processes in these systems stochastically bound the number of customer in the basic system both above and below. Then it is necessary to prove that the difference between these comparative processes is stochastically bounded. Moreover, the behavior of these processes is close to the one of the standard systems (f=1, a=1), and we can employ some previous results.
Key Words: multichannel queuing system, impatient customer, regenerative flow

On stable states of GMM clustering

Dvora Toledano-Kitai, Zeev Volkovich
Department of Software Engineering, ORT Braude College, Israel

The general clustering Gaussian Mixture Model (GMM) is a flexible and powerful probabilistic clustering tool. The technique typically employed to fit this model to given data is the famous expectation maximization EM algorithm, which is used to determine a maximum likelihood estimate of the model parameters. One essential issue in applying this algorithm is the number of components chosen. Choosing too many components could result in overfitting the data, whereas a mixture with too few components may not be sufficiently adaptable to approximate the true source model.
In the proposed approach for solving this model selection problem, the GMMs resulting from several tested numbers of components are compared by means of sample clustering of the underlying data and the perturbed version of the data. Thus, similarity between the partitions is measured by the *p*-values of the appropriate two-sample test statistic. Two main types of statistics are considered: The first corresponds to the "*k*-means" configuration of the spherical equally sized clusters. The second consists of identically shaped clusters. No simulation method intended to synchronize partitions in the compared samples is used. An empirical distribution of the *p*-values, which is less concentrated at origin,

supplies an estimation of the true number of components. The numerical experiments demonstrate the capabilities of the proposed methodology.
Key Words: Clusterin, Cluster Stability, GMM

The Minimum Pseudodistance Approach: an Application to Extreme Quantile Estimation in Finance

Aida Toma, Silvia Dedu
Department of Applied Mathematics, Bucharest Academy of Economic Studies, Romania

In this paper we study minimum pseudodistance estimators (MPE) for tail probabilities and quantiles of the generalized extreme value (GEV) and the generalized Pareto (GP) distributions. These distributions play an important role in quantile estimation based on extreme value theory, being applied in modeling financial risks in finance and insurance. The MPEs are defined by minimization of an empirical version of a pseudodistance between the assumed model and the true model underlying the data. The behavior of the estimators depends on a positive tuning parameter which controls the trade-off between robustness and efficiency. We discuss theoretical properties of the MPEs for the GEV and GP distributions, among them the consistency, asymptotic normality and robustness. The influence functions of the estimators are derived. The empirical performance of the proposed approach is also analyzed. In this respect, using Monte Carlo simulations, we assess the relative efficiency between a MPE and the MLE, for different values of the tuning parameter and various sample sizes. Then we study the performance of the MPEs for extreme quantile estimation using real financial data. The obtained results show that the MPEs outperform the MLE in the considered context.

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.
Key Words: Minimum pseudodistance estimators; Generalized extreme value distribution; Generalized Pareto distribution.

Stochastic Modelings in Software Reliability

Nuria Torrado
Center for Mathematics, University of Coimbra, Portugal

A software system is a repairable system, i.e., it can be restored to fully satisfactory performance by any method, other than replacement of the entire system, after failing to perform one or more of its functions satisfactorily. The demand for complex software systems has increased within the first decade of the 21st century. Therefore, the reliability of software systems has become a major concern for our modern society.

A software reliability model (SRM) is a mathematical tool to evaluate the software quantitatively. The SRM's have been extensively developed in the literature. Most SRM's are based on stochastic counting processes, such as binomial process, pure birth process and nonhomogeneous Poisson process (NHPP). One may refer to two excellent books by Lyu (1996) and Pham (2007) on this topic. These stochastic models attempt to model either the times between successive failures of a piece of software or the number of failures in fixed time periods.

Some interesting software reliability modeling developed in the last two decades can be found in Wiper (2007), Pievatolo et al. (2012) and Torrado et al. (2013). In this work, a short review of these references is given, and also, I will discuss recent results (and some ongoing work) obtained in collaboration with R.E. Lillo and M.P. Wiper.

References

1. Lyu, M.R.(1996), Handbook of Software Reliability Engineering, IEEE Computer Society Press, McGraw-Hill.
2. Pham, H. (2007). System Software Reliability. Springer.
3. Pievatolo, A., Ruggeri, F. and Soyer, R., A bayesian hidden markov model for imperfect debugging. Reliability Engineering and System Safety, 103, 12-21 (2012).
4. Torrado, N., Wiper, M.P. and Lillo, R.E., Software reliability modeling with software metrics data via gaussian processes. IEEE Transactions on Software Engineering, 39, 1179–1186 (2013).
5. Wiper, M.P., Software reliability: Bayesian analysis, in F. Ruggeri, R. Kennet and W. Faltin (editors) Encyclopedia of Statistics in Quality and Reliability. Wiley, New York (2007).

Branching Processes: Forecasting Human Population

Plamen I. Trayanov[1], Maroussia Slavtchova-Bojkova[2]
[1]Department of Probability, Operational Research and Statistics, Faculty of Mathematics and Informatics, Sofia University, [2]Department of Probability, Operational Research and Statistics, Faculty of Mathematics and Informatics, Sofia University, Institute of Mathematics and Informatics, Bulgaria

In this work we develop a new technique using the Crump-Mode-Jagers branching process theory to model human population. We are addressing questions like: How the population grows according to given scenarios and how these results could be used in decision making to choose an appropriate demographic policy. Nowadays, such issues are especially important in view of the tasks before the knowledge-based society.
Our aim is to estimate the growth of young population based on national survey data taking into account the internal and international migration.

Acknowledgement: This work is partially supported by the state funds allocated to the Sofia University, grant No 111/2013 and by funds of the National Project for supporting doctoral students and young scientists in Mathematics and Computer Science at Sofia University “St. Kliment Ohridski” and co-financed by the European Social Fund of EU, project No BG051PO001-3.3.06-0052.
Keywords: General Branching Process, Demography, Population projections, Malthusian parameter.

Multivariate Control Charts Based on Bayes Factors

Kostas Triantafyllopoulos[1], Sotiris Bersimis[2]
[1]School of Mathematics and Statistics, University of Sheffield, UK, [2]Department of Statistics and Insurance Science, University of Piraeus, Greece

Multivariate statistical process control (MSPC) has gained increased attention over the recent years, in particular regarding the simultaneous control of variables of non-industrial processes. Such processes frequently exhibit autocorrelation (serial correlation) over time. Local dynamics and in-control changing mean over time call for non-standard MSPC procedures to accommodate, model and describe the in-control structural state over time and to identify out-of-control process performance in Phase II. A common approach in such a problem is to fit in Phase I some suitable time-series model, capable of describing the dynamics of the multivariate process, and to define out-of-control signals issued by deviations from the distribution of the residuals (for example deviations from the mean and variance of the residuals). In this talk we

propose that such a control procedure can be carried out by comparing the residual distribution under the in-control state with the distribution obtained by considering certain deviations from the in-control distribution. This process can be implemented by considering Bayes factors, which are similar to sequential likelihood ratios, but have certain advantages, in particular for online prediction and hence more suitable for Phase II analysis. We describe this approach and we discuss deviations of the mean and dispersion from the in-control residuals. A state-space model is chosen to describe the process dynamics and to form the Bayes factors sequentially over time. Non-parametric control charts are then employed to decide whether the distance of the two distributions is significant, hence to identify out-of-control behaviour. The proposed approach will be evaluated considering artificially simulated data and comparisons with standard multivariate control charts. The methodology will be illustrated with real data.

Keywords: Multivariate control charts, Autocorrelated process, Bayes factor, State Space model, Non-parametric control charts

Modeling the relationship between temperature and daily mortality in Cyprus

Haritini Tsangari[1], Zoi Konsoula[1], Stephanie Christou[1], Kyriacos Georgiou[2], Edna Yamasaki[1]

[1]University of Nicosia Research Foundation, [2]Cyprus Center for European and International Affairs, University of Nicosia, Cyprus

Climatic changes, such as large temperature fluctuations and increase in the occurrence of heat waves, have been evidenced to affect mortality worldwide. In this paper we examine the effect of high temperatures on mortality in Cyprus, an island which is characterized by a Mediterranean climate. The modeling approach is described. First, the temperature function is created within the newly-developed framework of distributed lag non-linear models, to simultaneously capture non-linearities and delayed effects. The temperature function is, then, incorporated in a Generalized Linear Model with a quasi-Poisson distribution to allow for overdispersion, together with possible confounders such as meteorological indicators, trends and seasonality. Comparisons are additionally made, regarding the effect of temperature on mortality, between inland and coastal areas. All the results are presented in a tabular or graphical form and the conclusions are discussed.

Keywords: heat waves, mortality, distributed lag non-linear model, strata constraints, hot threshold, GLM, quasi Poisson, harvesting effect.

AN ADJUSTED NETWORK INFORMATION CRITERION FOR MODEL SELECTION IN STATISTICAL NEURAL NETWORK MODELS

Christopher Godwin Udomboso, Angela Unna Chukwu, Isaac Kwame Dontwi
University of Ibadan, Nigeria

In selecting a Statistical Neural Network model, the Network Information Criterion has been observed to be sample biased, because it does not account for sample sizes. The selection of a model from a set of fitted candidate models requires objective data-driven criteria. In this paper, we derived and investigated the ANIC criterion, based on Kullback's symmetric divergence, which has been designed to be an asymptotically unbiased estimator of the expected Kullback-Leibler information of a fitted model. The analyses show that on a general note, the ANIC improves model selection in more sample sizes than does the NIC.
Key Words: Statistical neural network, network information criterion, adjusted network information criterion, transfer function

Combinatorial Approach to Statistical Design of Experiment

Petya Valcheva
Sofia State University "St. Kliment Ohridski"

The paper describes the various types of combinatorial designs applied in Statistical Design of Experiment. We give more detailed information about Balanced Incomplete Block Designs (BIBDs) and Orthogonal Arrays (*OAs*). The "difference matrix"-method for achievement of OAs is also described. Based on it two new constructions of orthogonal arrays (6,15) and (6,20) are obtained. They are non-isomorphic to the already known constructions of such arrays. The known (7,12)-OA is also discussed.
This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3:3:06-0052(2012/2014).
Keywords: difference matrix, balanced incomplete block design, quasi-difference matrix, orthogonal array.

Markovian Analysis of a three echelon supply chain with stochastic demand, lost sales, (S, s) continuous review policies and Coxian 2-phase lead times

Vidalis Michael, Papadopoulos Chryssoleon, Vrisagotis Vassilios
University of the Aegean, Department of Business Administration, Greece

Inventory control plays an important role in supply chain management. Decisions regarding order quantity and reorder point can satisfy customers' demands, reduce the operation costs and increase profitability. This study investigates a three echelon supply chain consisting of a retailer a wholesaler and a manufacturer. Demand and lead times are uncertain. The wholesaler and the retailer follow continuous review inventory policies (Si, si), i=1,2. The lead times between successive nodes are random and follow the Coxian-2 distribution. Assuming lost sales at the retailer and infinity capacity at the manufacturer, we try to explore the performance of the supply chain system. The system is modeled as a continuous time Markov process with discrete space. The structure of the transition matrices of these specific systems is examined (block partitioned) and a computational algorithm is developed to generate them for different values of system characteristics. Based on the transition probabilities and steady-state behavior of the system the long-run behavior of the inventory model is analyzed. The proposed algorithm allows the calculation of performance measures –fill rate, cycle times, average inventory (WIP). Moreover, expressions for the holding costs and shortage costs are derived.
The main contribution of this work is the presentation of an exact evaluative model to calculate performance measures of the system in the case of batch production with Coxian replenishment times. This model can be used as a generative one to determine the values of the parameters that optimize the behavior of the system given an objective function.
Key Words: Three-echelon supply chain, performance measures, Markov analysis, lost sales, continuous review (S,s) inventory policies, Coxian-2 Lead times

ACKNOWLEDGEMENTS
The authors Michael Vidalis and Papadopoulos Chrysoleon have been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

On 21 century's misusing of the classical Pearson's goodness-of-fit test

V. Voinov, R. Makarov
KIMEP University, Kazakhstan

When testing for multivariate normality the Cardoso De Oliveira & Ferreira's test, and power divergence statistics cannot follow in the limit the chi-squared distribution with *M*-1 degrees of freedom, where *M* is the number of equiprobable grouping cells, as it was announced in two recent articles in 2010 and 2013. The exact limiting distribution of the Cardoso De Oliveira and Ferreira's test has been derived. We show that, even if the above tests will be implemented correctly, their power with respect to alternatives close to multivariate normal distribution is much less than that for other well known by the date tests for multivariate normality. We show that the above tests as well as the classical Pearson's sum that use estimates of unknown parameters based on raw data do not reject the null hypothesis as often as it should be. An extensive simulation study and two classical data sets were used to illustrate the theoretical arguments.
Keywords: Pearson's sum; power divergence tests; estimates based on raw data; multivariate normality; power of tests.

A fuzzy trend model for multivariate time series with seasonal components

Erika Watanabe, Norio Watanabe
Chuo University, Japan

Analysing trend for multivariate time series is an important issue. A fuzzy trend model has been proposed for estimating trends in multivariate time series. This fuzzy trend model can decompose trends into common and individual trends. However, seasonality is not considered in this model. In this study we propose a model including seasonality. Another problem of the former model is that common trend might differ from each series when there is big difference among series. Therefore we propose a scaling model which can decompose trends effectively by introducing weights. Usability of proposed models is demonstrated by simulation studies and applications.
Key Words: seasonality common trend, decomposition of trend, scaling

Some Fuzzy Correlation Coefficients for Bivariate Fuzzy Data

Norio Watanabe[1], Erika Watanabe[2]
[1]Chuo University, [2]Graduate School of Chuo University, Japan

Several fuzzy correlation coefficients have been proposed for bivariate fuzzy data. However, it is not clear which correlation coefficient should be used. From the practical viewpoint it is important to select the appropriate correlation coefficient for each data in order to extract information on correlation, since fuzzy data involves a diversity of fuzziness. Therefore we introduce two fuzzy correlation coefficients in addition to the one, which has been proposed for a practical use. Moreover we clarify what kind of fuzzy data is suitable for each fuzzy correlation coefficient. Usability of the proposed method is demonstrated by applying to real fuzzy data on sensitivity, that is, Kansei data.
Keywords: Correlation analysis, Kansei data

Testing unit root test based on Polyvariogram

WU, Ka Ho
The Chinese University of Hong Kong, Hong Kong

Unit root tests have been one of the main research areas in time series analysis for several decades.
Unfortunately, the low power and/or size distortion in some difficult situations are still the problems needed to be solved so that software can be developed for practical use.
In this talk, we present solid and powerful test procedures with theoretical support using the statistics based on polyvariograms for identifying the order of differencing d for a given time series.
The proposed test methods are shown to be more powerful and performs well in difficult situations. The asymptotic distribution of the test statistic is derived. Simulation studies demonstrate the superiority of the proposed method to some mostly used methods in the literature.
Key Words: ARIMA models; asymptotic distribution; power of tests.

Ranking charity applications

Cong Xu, James M Freeman
Manchester Business School, The University of Manchester, UK

How can a charity ensure that only its most worthy causes are supported? One approach that has proven effective in the past is AHP (Analytic Hierarchy Process). Now a new method – based on a hybrid

AHP / Evidential Reasoning (ER) adaptation - has become available, claiming distinct advantages over straight AHP. By contrasting the two procedures for a real-life dataset – we demonstrate AHP/ER's superiority in both theoretical and practical respects.
Keywords: Analytic Hierarchy Process, Dummy variables, Evidential Reasoning, Intelligent Decision Software, Utility

Limit Theorems for Supercritical Branching Random Walks with Heavy Tails

Elena Yarovaya
Lomonosov Moscow State University, Russia

Continuous-time branching random walks on multidimensional lattices give an important example of stochastic multicomponent systems in which the evolutionary processes depend on the structure of a medium and the spatial dynamics. The structure of a medium is defined by the offspring reproduction law at the origin. The dynamics of such processes is usually described in terms of birth, death, and walks of particles on the lattice. Such a description allows to find various realistic applications of branching random walks. One of the principal problems in branching random walk models is study of the evolution of populations of particles at an arbitrary point of the lattice and on the entire lattice. The spectrum structure of an evolutionary operator of mean numbers of particles is of importance in analysis of properties of continuous-time branching random walks. Main results here are the limit theorems for supercritical branching random walks under the assumption that underlying random walk has a heavy tail.
This work was supported by RFBR grant 13-01-00653.
Key Words: Branching Random Walks, Green's Functions, Supercritical case, Limit theorems

On the Accuracy of the Risk Estimators

Georgios C. Zachos*
University of the Aegean, Department of Statistics and Actuarial – Financial Mathematics, Greece

The purpose of this paper is to evaluate the accuracy of the beta estimations that are suggested to be free of intervalling effect bias. To this end we examine the accuracy of the asymptotic estimators of betas by comparing them to OLS assessments as well as to beta estimations adjusted according to their tendency to regress towards one. Furthermore we employ above measurement for different intervals

among data observations and we re-examine our findings by taking into account the Corhay effect. In addition we use models that take into account Heteroskedasticity in residuals and we perform the same comparisons.
Keywords: Adjusted Risk Coefficients, Intervalling Effect Bias, Asymptotic Beta
JEL classification: C22, G12, G14

*Georgios C. Zachos is indebted to D. Konstantinides for his invaluable advices, E. Lamprou for providing datasets and D. Kokonakis and E. Papanagiotou for their valuable help.

A method for calculating life tables using archive data. An example from mountainous Rhodopi.

Konstantinos N. Zafeiris
Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, Greece

Using archive data, the Pomaks of Organi and Kehros (Greek Thrace) were studied. In order for the mortality transition to be evaluated a life table analysis was carried out. Results suggest a rapid mortality transition. Finally, Pomaks have converged to the Thracian population.
Keywords: Mortality, life tables, Pomaks, Thrace.

Demographic and Health Indicators in the Pomaks

Konstantinos N. Zafeiris[1], Christos H Skiadas[2]
[1]Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, [2]ManLab, Department of Production Engineering and Management, Technical University of Crete, Greece

Based on the Health State Theory several demographic and health state indicators were calculated for the Pomaks, a Slavic speaking population from Mountainous Rhodopi (Greece). Results indicate a rapid health transition and a general improvement of the health status of the population. Gradually, by the 1990s Pomaks converge to the total population of Thrace.
Keywords: Pomaks, Thrace, Health State Indicators

Applications of the gambling team method to the solutions of problems deal with occurrences of words in random texts

Krzysztof Zajkowski
Institute of Mathematics, University of Bialystok, Poland

We show applications of the gambling team technique, invented by Li, to investigations of the expected waiting time until one of a given collection of patterns is observed and probabilities for each pattern to be the first to appear in sequences of independent and Markov dependent trials.
Key Words: martingales, stopping times, optional stopping theorem, gambling team technique, Markov chains
Bibliography
E. Fisher, S. Cui, Patterns generated by mth-order Markov chains, Statist. Probab. Lett. 80 (2010), 1157-1166.
H. Gerber, S-Y.R. Li, The occurrence of sequence of patterns in repeated experiments and hitting times in a Markov chain, Stochastic Processes and Their Applications 11 (1981), 101-108.
J. Glaz, M. Kulldorff, V. Pozdnyakov, J.M. Steele, Gambling teams and waiting times for patterns in two-state Markov chains, Journal of Applied Probability 43 (2006), 127-140.
S-Y.R. Li, A martingale approach to the study of occurrence of sequence patterns in repeated experiments, The Annals of Probability, Vol. 8. (1980), 1171-1176.
V. Pozdnyakov, On occurrence of patterns in Markov chains: method of gambling teams, Statistics and Probability Letters 78 (2008), 2762-2767.
K. Zajkowski, A note on the gambling team method, Statistics and Probability Letters 85 (2014), 45-50. doi: 10.1016/j.spl.2013.11.003

On Improving The Maximum Likelihood Estimators

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

Let a sample of size n be drawn from an absolutely continuous two-parametric distribution, where the first parameter is that of interest, while the second one is nuisance. We consider only the case when the nuisance parameter is either scale or location. Our goal is to estimate the parameter of interest.
The well-known approach to the problem is that based on the maximum likelihood principle. Together with the maximum likelihood estimator we investigate several modifications, e.g., the bias corrected estimator or the maximum integrated likelihood estimator. Under some usual

regularity conditions on the distribution, we study the second order properties of all estimators and concentrate on comparison of their biases and mean square errors. Since, in general, the exact distributions of the estimators are not available due to non-explicit forms of these estimators, we are focused on the asymptotic comparison.
Then we also try to improve all the estimators using the methodology given in the paper of Ghosh and Sinha (1981).
As the examples of results obtained, different two-parametric families of distributions are considered. The results are supported with numerical calculations.

References

1. Ghosh, J. K. and Sinha, B.K., A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. Ann. Statist. 9, 1334–1338 (1981).

Strong Approximation of the Random Sums with Applications in Queuing and Risk Theories

Nadiia Zinchenko
Nizhyn State Mukola Gogol University, Ukraine

We present some general results concerning sufficient conditions for strong approximation of random sums

$$S(N(t)) = \sum_{i=1}^{N(t)} x_i$$

by a Wiener or α-stable Lévy process under various conditions on the counting renewal process N(t) and random summands $\{x_i, i \geq 1\}$. Corresponding proofs are based on the rather general theorems about strong approximation of the superposition of càd-làg processes (Zinchenko (2013)). This topic is interesting not only from theoretical point of view but also due to numerous practical applications, since mentioned random sums often appear in useful applications in the queuing theory (accumulated workload input into queuing system during time interval (0, t)), in the risk theory (total claim amount to insurance company up to time t), in the financial mathematics (total market price change up to time t) and in certain statistical procedures. As a consequence a number of limit theorems for risk processes in classical Cramer-Lundberg and renewal Sparre Andersen risk models can be obtained, particularly, strong and weak invariance principle for risk processes, diffusion and stable approximation of ruin probabilities, various modifications of the LIL and Erdös-Rényi-Csörgő-Révész-type SLLN for risk processes, which describe the rate of growth and fluctuations of mentioned processes and are useful for planning the

insurance activities and reserves. Similarly the case of risk models with stochastic premiums is investigated as well as certain queuing models.
Keywords: Random sums, Strong approximation, Invariance principle, Strong limit theorems, Risk process, Queuing models

References

1. Zinchenko, N. Almost sure approximation of the superposition of the random processes, Methodology and Computing in Applied Probability, http://dx.doi.org/10.1007/s11009-013-9350-y (2013).

Approximate Bayes estimation of the parameters of gamma exponentiated exponential model

Nabil Zougab
Laboratory LAMOS, University of Bejaia, and University of Tizi-Ouzou, Algeria

Recently, the gamma-exponentiated exponential distribution is proposed and studied in detail. The method of maximum likelihood (MLE) is used for estimating the three parameters of this model. In this paper, we propose the Bayesian alternative procedure to parameters estimation in the gamma-exponentiated exponential model. Monte Carlo simulations are performed to observe the performance of the proposed method.
Key Words: Bayesian procedure; gamma-exponentiated exponential; Maximum likelihood; Monte Carlo.

Estimating the Right Endpoint in the Weibull Max-Domain of Attraction

M. Isabel Fraga Alves
Universidade de Lisboa, FCUL, DEIO and CEAUL, Lisboa, Portugal

A new semi-parametric method for estimating the right endpoint of a distribution is presented, when the distribution function belongs to the Weibull max-domain of attraction. Although the proposed estimator has been built on properties of distributions with finite right endpoint in Gumbel domain (Fraga Alves and Neves (2013), the same estimator (FAN) it is also a consistent estimator for distributions lying in the Weibull max-domain of attraction and its good performance is illustrated on some finite sample situations. The proposed estimator for the right endpoint returns values always larger than sample maximum and do not need the estimation of any extra parameter. This constitutes a major advantage in comparison to the available semi-parametric endpoint estimators, for which the extrapolation beyond the sample range is not guaranteed. This inadequacy of the existing estimators often leads to some disappointing results in practical applications, with estimates yields that may be lower than the observed maximum from the data. For sake of comparison, the exact performance FAN estimator is compared with naive maximum

estimator (MAX), and with endpoint moment estimator (MOM) from Dekkers et al. (1989)). The best exact performance of FAN estimator is attained for values of extreme value index close to zero, and can be seen as an alternative to MOM estimator, which reveals large variability in the top sample, and a complementary inferential method to MAX estimator, which underestimates the true value for the right endpoint.
This is a joint work with Cláudia Neves, University of Aveiro, Portugal.

References

1. Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L., A moment estimator for the index of an extreme-value distribution. Annals of Statistics, 17:4:1419–1917 (1989).
2. Fraga Alves, M. I. and Neves, C., Estimation of the finite right endpoint in the Gumbel domain. Statistica Sinica: Preprint doi:10.5705/ss.2013.183 . (2013).

A Probabilistic Framework for the Evaluation of the Environmental Performance in Wine Industry

D. Dede[1], S. Bersimis[2], and D. Georgalelos[3]
[1]Quality Assurance Program, Open University, Greece
[2]Dept.of Statistics and Insurance Science, University of Piraeus, Greece
[3]Department of Business Administration, University of Piraeus, Greece

Life-cycle analysis is a scientific method that is related to sustainable consumption and production. This technique is based on a large amount of multivariate data that are difficult to comprehend and to interpret. Bersimis and Georgakelos (2013) used the method of Life-cycle analysis in conjunction to principal component analysis and they provided a graphical method in order to interpret the available multivariate data. In this work, we appropriately modify the method proposed by Bersimis and Georgakelos (2013) in order to analyze a set of Life-cycle analysis data that was acquired from wine industry of Greece. As was revealed by this application, the method proposed by Bersimis and Georgakelos (2013) is proved to be robust and useful independently of the data source.