



HAL
open science

Wrapper aggregation for classification

Tristan Mary-Huard

► **To cite this version:**

Tristan Mary-Huard. Wrapper aggregation for classification. ASMDA 2009 : 13. International Conference, Jun 2009, Vilnius, Lithuania. pp.47. <hal-01197558>

HAL Id: hal-01197558

<https://hal.science/hal-01197558v1>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

XIII International Conference

**“Applied Stochastic Models and Data Analysis”
(ASMDA-2009)**

ABSTRACTS

June 30-July 3, 2009,

Vilnius, Lithuania

The XIII International Conference APPLIED STOCHASTIC MODELS AND DATA ANALYSIS

(June 30-July 3, 2009, Vilnius, Lithuania)

Organised by

- Applied Stochastic Models and Data Analysis International Society (ASMDA);
- Vilnius Gediminas Technical University, Vilnius, Lithuania;
- Institute of Mathematics and Informatics, Vilnius, Lithuania;
- Lithuanian Operational Research Society (LitORS);

In cooperation with

- Vilnius University, Vilnius, Lithuania;

Supported by

- Lithuanian State Foundation of Study and Research

Chairman of the Conference:

Prof. Jacques Janssen
ENRIA (Brest) and JACAN (Tubize), France

Chairman of the Programme Committee:

Prof. Jacques Janssen
ENRIA (Brest) and JACAN (Tubize), France

Chairman of the Organising Committee:

Prof. Leonidas Sakalauskas
Institute of Mathematics and Informatics,
Lithuania

INTERNATIONAL SCIENTIFIC PROGRAMME COMMITTEE

S. Aivazian, Russia
A. Andronov, Latvia
V. Bagdonavitchius, Lithuania
A. Bagirov, Australia
N. Balakrishnan, Canada
M. Bardos, France
J. Blanchet, USA
W. K. Brauers, Belgium
M. Brown, USA
E. Bulinskaya, Russia
P. S. Chan, Hong Kong
E. Cinlar, USA
D. Donchev, Bulgaria
G. Dzemyda, Lithuania
R. Elliott, Canada
S. Ermakov, Russia
N. Farmakis, Greece
Y. Guermeur, France
J. Glaz, USA
A. Guénoche, France
F. Guillet, France
G. Govaert, France
S. Haberman, UK
J. Janssen, **Chair**
D. Ho, France
A. Karagrighoriou, Cyprus
Y. Kharin, Belarus
T. Kollo, Estonia
E. Kontoghiorghes, UK
V. Koroliuk, Ukraine
S. Lallich, France
L. Lebart, France
Mei-Ling Ting Lee, USA
R. Leipus, Lithuania
B. Lemeshko, Russia
P. Lenca, France
N. Limnios, France
Bo Lindqvist, Norway
G. MacKenzie, Ireland
R. Manca, Italy
L. Markus, Hungary
S. McClean, Northern Ireland
K. Mosler, Germany
M. Nikulin, France
I. Nourdin, France
J. Ott, USA
J. Owsinski, Poland
J. Pagès, France
T. Papaioannou, Greece
S.M. Ross, USA
L. Sakalauskas Lithuania
G. Saporta France
L. Saulis, Lithuania
D. Silvestrof, Sweden
P. Sirirangsi, Thailand
C. H. Skiadas, Greece
M. J. Valderrama, Spain
P. Vasileiou, Greece
V. Voinov, Kazakhstan
A.E.Xavier, Brasil
S. Zacks, USA
E. Zavadskas, Lithuania
K. Zografos, Greece

Organising Committee

Vilijandas Bagdonavichius
Vaida Bartkute-Norkuniene
Daiva Kackauskaite
Olga Kurasova
Remigijus Leipus
Virginijus Marcinkevičius
Leonidas Sakalauskas - **Chair**
Valentinas Skarzhauskas
Christos Skiadas
Ana Uspuriene
Ingrida Vaiciulyte

THE ANALYSES OF INDIVIDUAL PATIENT PATHWAYS: INVESTIGATING REGIONAL VARIATION IN COPD READMISSIONS

Shola ADEYEMI, Eren DEMIR, Thierry CHAUSSALET
Health and Social Care Modelling Group, University of Westminster, London, UK
E-mail: adeyems@wmin.ac.uk

Regional variation for those treated for chronic obstructive pulmonary disease (COPD) has been studied amongst many public health issues. However, for those experiencing frequent readmissions were not considered. The objective is to determine the probability of a patient being in the high risk group in their next readmission, given patient and regional specific previous readmissions, i.e. capturing individual clinical pathways of patient's leading to multiple readmissions. Data on COPD patients were extracted from the national (England) hospital episodes statistics dataset. Individual patient pathways from initial admission through to more than four readmissions are captured. Here, patient's are classified into the low risk group (LRG) and the high risk group (HRG) of readmission. This classification is based on the definition of readmission, where we have previously developed a stochastic modelling approach to determine the appropriate width of a time window in defining readmission. Hence, using this classification as a response variable, a random effects continuation-ratio logit model is used to examine regional disparities in progressing through multiple readmissions. Significant regional differences were found in each readmission and classification groups (HRG or LRG). For example, COPD patients from the North West of England were found to be 16% more likely to experience multiple readmissions than Yorkshire and The Humber region.

STOCHASTIC MODELS OF TRANSPORT FLOWS

Larisa AFANASEVA, Ekaterina BULINSKAYA
Moscow State University, Russia
E-mail: ebulinsk@mech.math.msu.su

This research continues the investigation carried out by authors in [1]. We begin by treating one-lane highway with long distances between traffic lights. That means, we do not need to take into account the cars size. Under such assumption we study the delays caused by traffic lights and cars flow density. The sufficient conditions for existence of steady-state are established along with optimization of traffic lights performance.

Next we investigate the case of short distance between the traffic lights where the vehicles size plays a crucial role in the study of traffic congestion.

We use the mean field techniques as well as renewal, Markov and regenerative processes.

ON THE CONFIDENCE INTERVALS FOR MEAN RESPONSE IN EXPERIMENTS WITH MIXTURES

Kadri Ulas AKAY
University of Istanbul, Turkey
E-mail: kulas@istanbul.edu.tr

Confidence intervals for estimated mean response provide valuable information about the usefulness of regression models. For example, the interval estimate gives an indication of how much uncertainty there is in our estimate of the true mean. Confidence interval for mean response in each experimental run is obtained by multiple regression analysis. In experiments with mixture, the presence of natural constraints or/and some additional linear constraints between the components (or mixture variables) require the use of specific multiple regression models such as Scheffe canonical polynomials for the modeling of the mixture system. In this study, as an alternative to obtain the confidence intervals in each experimental run for the mean response, a graphical approach has been proposed by using the constraints between the components to compare the models and choose the appropriate model for the mixture system, depending on the confidence intervals for the mean response. The proposed graphical approach is examined on an example in the literature.

DYNAMIC RELIABILITY AND UNCERTAINTY ANALYSIS OF DELAYED HYBRID SYSTEM

Robertas ALZBUTAS
Lithuanian Power Institute, Lithuania
robertas@mail.lei.lt

Recently the Stimulus Driven Theory of Probabilistic Dynamics (SDTPD) or the Theory of Stimulated Dynamics (TSD) is developed for analytical modelling and simulation of hybrid (continuous-discrete) systems as well as for dynamic reliability considerations. The theory deals with the concept of stimulus and how it can be implemented. In addition, the focus can be set on the delays considerations. Because stimuli activation and random delays conditions the events, the history of stimuli activations during the accident transients do matter in calculating the frequencies, and extensions of the Markov process equations accounting for these features is considered. A great number of dependencies and probabilities to evaluate are the most important limitation of using Markov techniques for large sets of system components. Thus, development and grouping of sequence becomes necessary and some general and test case grouping applications are investigated. A non-Markovian treatment is provided in order to adapt SDTPD for practical applications, mostly in the context of delayed system reliability analysis. The development of it as well as related methods for uncertainty analysis have been used as a basis for test case simulation in the perspective of it applications for severe accident scenarios and uncertainty analysis. Since the application of TSD related to delayed systems reliability concept needs a formal approach, the further consideration of test case simulation and additional investigation related to sequence grouping and parameters uncertainty influence to the result uncertainty is performed. The analysis of sequence of delayed dynamics and results sensitivity to the uncertain parameters is introduced as a possible way for analysis of TSD based simulation results. The issues of delayed systems reliability and timing importance are discussed in relation to the test case results.

The considered analysis of delayed system reliability and uncertainty clearly allows a systematic generation and analysis of random parameters and timing. The developed approach for delayed system simulation can efficiently use the available computer resources and at same time to allow collect information for analysis considering uncertainty, i.e. simulate delayed system result under consideration and represent uncertainty of certain result.

THE INTERPLAY BETWEEN FINANCIAL AND DEMOGRAPHIC RISKS IN A PENSION ANNUITY SYSTEM

Valeria D'AMATO, Emilia Di LORENZO, Maria RUSSOLILLO, Marilena SIBILLO
University of Salerno, Italy
E-mail: diloremi@unina.it

The financial risk in pension annuity portfolios constitutes the main uncertainty source. The paper deals with the estimate of the impact of the financial risk within a stochastic surplus analysis, according to several demographic scenarios. Risk measures for a homogeneous annuity portfolio are provided, by means of formulas apt to take into account the interactions between the main risk components, both financial and demographic. Numerical examples clarify the results.

EFFICIENT BOOTSTRAP APPLIED TO THE POISSON LOG-BILINEAR LEE CARTER MODEL

Valeria D'AMATO, Steven HABERMAN, Maria RUSSOLILLO
University of Salerno, Italy
E-mail: Valeria.Damato.1@city.ac.uk

In life insurance, the companies have dealt with two fundamental types of risks when issuing contracts: financial risk and demographic risk. As regards the latter, the uncertainty of the future developments of the mortality is of particular importance to financial institutions involved with selling annuities and pensions-related products. Recent work has focused on modelling the trend in mortality as a stochastic process. The most popular method for modelling the death rates is the Lee Carter model (1992), because it is easy to implement and outperforms other models with respect to its prediction errors (e.g.,

Koissi et al., 2006; Melnikov and Romaniuk, 2006). This methodology has become widely used and there have been various extensions and modifications proposed to attain a broader interpretation and to capture the main features of the dynamics of the mortality intensity, e.g. Booth, Maindonald and Smith, 2002 Hyndman and Ullah 2005, Renshaw and Haberman 2003. In this research, we propose a procedure for obtaining mortality projections in a robust way, on the basis of a log bilinear Poisson Lee Carter model (Renshaw and Haberman 2003). In that respect, we provide confidence intervals for forecasted quantities derived by using simulation techniques: these are important because of the non-linear nature of the quantities under consideration. Different simulation strategies have been applied in the log-bilinear Poisson setting: the Monte Carlo approach and the Bootstrap procedures. In a comparative study of the simulation strategies under consideration, Renshaw and Haberman (2008) highlight that the Monte Carlo simulation approach should not be used for risk measurement in LC modelling, since different choices of constraints result in widely differing simulated confidence and prediction interval widths. On the basis of these findings, we consider the bootstrap simulation approach to measure the uncertainty affecting the mortality projections. As regards Poisson LC modelling, the bootstrap procedure can be accomplished in different ways. In particular, four alternative approaches have been suggested: the semi parametric bootstrap from the Poisson distribution (Brouhns et al. 2005), the semi parametric from the multinomial distribution (Brouhns et al. 2005), the residual bootstrap proposed by Koissi et al (2006) and the variant of the latter illustrated in Renshaw and Haberman (2008). All of these methodologies generate bootstrap samples of the death rates and the Poisson model is re-estimated with each of these simulated data matrices. In this paper, we take into account the semi parametric bootstrap from the Poisson distribution. In particular, we intend to make efficient the bootstrap procedure by using a set of techniques, the so-called variance reduction techniques (VRTs). Each of these can be successfully applied to a particular class of problems in order to reduce the variance of the simulation estimate. All are based on the idea of obtaining an acceptable accuracy of the estimate by solving the problem of reducing the variance of the estimator. Finally, we provide an application of the proposed simulation procedure to the demographic and actuarial field. The empirical results are presented using a range of graphical analyses. We stress the financial impact of the results for a homogeneous portfolio of pension annuities.

References

1. Brouhns, N., Denuit, M., van Keilegom, I. 2005. Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 3, 212-224.
2. Booth, H., Maindonald, J., and Smith, L. 2002. Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56, 325-336.
3. Efron, B. Tibshirani, R.J. 1993. *An introduction to the bootstrap*. Chapman & Hall, New York & London.
4. Hyndman, R.J., and Ullah, S. 2005. Robust forecasting of mortality and fertility rates: a functional data approach. Working paper, Department of Econometrics and Business Statistics, Monash University. <http://www-personal.buseco.monash.edu.au/~hyndman/papers/funcfor.htm>.
5. Lee, R.D., Carter, L.R. 1992. Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87, 659-671.
6. Renshaw, A.E., Haberman, S. 2003a. Lee-Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections. *Applied Statistics*, 52, 119-137.
7. Renshaw, A.E., Haberman, S. 2003b. On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, 32, 379-401.
8. Renshaw, A.E., Haberman, S. 2003c. Lee-Carter mortality forecasting with age specific enhancement. *Insurance: Mathematics and Economics* 33, 255-272.
9. Renshaw, A.E., Haberman, S. 2008. On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics and Economics*, doi:10.1016/j.insmatheco.2007.08.009
10. Ross S. 2002. *Simulation*. Hartcourt Academic Press.

A NON-HOMOGENEOUS SEMI-MARKOV DISABILITY INSURANCE MODELS.

Guglielmo D'AMICO, Montserrat GUILLEN, Raimondo MANCA
University G. D'Annunzio of Chieti-Pescara, Italy
E-mail: raimondo.manca@uniroma1.it

Abstract. In this paper a stochastic model for disability insurance contracts is presented. The model is based on a discrete time non-homogeneous semi-Markov process to which the backward recurrence time process is joined. This permits to study in a more complete way the disability evolution and to face in a more effective way the duration problem. To describe the time evolution of claims and benefits we introduce a reward structure. In this way we construct duration dependent indicators on which it is possible to assess the fairness of the contract. The model is applied to a sample of contracts drawn at random from a mutual insurance company.

ON NONLINEAR REGRESSION MODEL FOR CORRESPONDENCE MATRIX OF TRANSPORT NETWORK

Alexander ANDRONOV, Diana SANTALOVA
Riga Technical University, Latvia
E-mail: diana.santalova@rtu.lv

A nonlinear regression model for a forecasting of passenger flow between various geographical points (towns) is described. Unknown parameters are estimated using aggregated data when only the information on the departed from each town passenger's quantity is available. As the estimation efficiency criterion the weighted sum of residual squares is used.

AN ANALYSIS REGARDING THE IMPACT OF SHOCKS TO CAPITAL ON A SOFT INSURANCE MARKET

Raluca Meda ANTAL, Gabriela Maria MIHALCA
"Petru Maior" University, Romania
E-mail: raluca4c@yahoo.com

In generating a viable starting point for our study we mention the main theoretical contributions to the specialized literature regarding the relationship between cycles and insurer capital such as Winter (1988, 1991, 1994), Gron (1994a), Cummins and Danzon (1997), Cagle and Harrington (1995), and Doherty and Garven (1995). We consider that the central themes of these papers are similar: shocks to capital can cause price increases and quantity reduction consistent with a hard market.

We believe that through an additional empirical study on the Romanian insurance market, we would provide evidence of the duration of any shock-induced price increases and whether costly external capital can also explain soft markets.

DYNAMIC-ENTROPY MONITORING DAILY STATE HUMAN HEART RATES

Valery ANTONOV, Janush DANILEVICH, Anatoly KOVALENKO, Artem ZAGAJNOV
St. Petersburg State Polytechnic University, Russia
E-mail: antonovvi@mail.ru

In the article results of modern researches holter monitoring of the electrocardiograms using methods of the nonlinear analysis are submitted. It is shown that fractal-entropy characteristics of a biological system condition and dynamics of their change depending on external and internal factors for urgent cardiology are of special interest. The settlement - experimental data of determined - chaotic parameters within day for patients of various ages are resulted and laws in their behavior are shown.

As initial entrance signals the full electrocardiograms of the patients which have been taken off with various frequencies of digitization (100, 200 and 500 Hz) are used. On the basis of the patented technology of a functional condition of an organism forecasting (the patent №2226067 Russian Federations) are received trends of determined - chaotic parameters such as correlation dimension, Shannon entropy, etc. within day for records holter monitors. Research of more than 20 patients of various age and a condition, taking place under the general anesthesia is made during various operations, thus their influence on the general condition of biological system is shown. The case of a lethal outcome at which the trend of correlation dimension some time (~ 2 hours) remained practically to constants is considered, and further his sharp recession up to zero (ventricular fibrillation) was observed.

PROBABILISTIC FORECAST FOR OUTSTANDING CLAIMS LIABILITIES

Miklos ARATO

Etvos Lorand University, Budapest, Hungary

E-mail: aratonm@ludens.elte.hu

For an insurance company it is crucial to determine the loss reserve. We regard stochastic reserving methods (see England and Verrall, 2002) as the probabilistic forecasts for outstanding claims liabilities. In our work we make a comparison of reserving methods using bootstrap methodology and scoring rules (see Gneiting and Raftery, 2006). Our bootstrap methodology is not a residuals bootstrap using aggregated data (run-off triangle). We use individual data set of an insurance company (more than 500000 policies and 100000 claims at the period 2000-2007). The policies are resampled with replacement and using claim history (years 2000-2003) of this policies we predict outstanding claims with different methods and compare these forecasts with the run-off result (years 2003-2007). We repeat this procedure 100 times. We analyze the differences between methods and the influence of the run-off triangle's period (monthly, quarterly and yearly) on the reserve.

BAYESIAN RELATIVE SENSITIVITY IN A REAL CASE OF REPLACEMENT PRIORITIES

J.P. ARIAS-NICOLES

University of Extremadura, Spain

E-mail: jparias@unex.es

Knowledge management is a key issue in the industrial field because of the difficulties in both properly quantifying experts' judgements and integrating them with available historical data. We discuss some sensitivity issues in dealing with experts' judgements in a real case study about strategies for the preventive maintenance of low pressure cast iron pipelines in a urban gas distribution network. We are interested in replacement priorities, as determined by the failure rates of pipelines deployed under different conditions. This allow us to detect changes in ranking of the optimal decisions when there are perturbations in the prior which model experts' judgements. The set of nondominated actions is considered a first approximation of the solution. However, it often is too big to take it as the solution of the decision problem. So, they are considered different criterions to choose an optimal alternative inside the nondominated set. Some authors recommend choosing the conditional Gamma-minimax or the posterior regret Gamma-minimax alternative. These alternatives often produce a relative increasing in posterior expected loss with respect to the optimal one, so we deal with sensitivity of ranking under changes in the prior especification. We recommend choosing the least sensitive alternative, which minimizes the sensitivity measure given by Arias-Nicoles et al. in 2008. Here, we compute the least sensitive alternative when the priors varies over some of the most usual classes in the literature, the generalized moments and e-contaminated classes.

MODIFIED TWO-STAGE LEAST SQUARES METHOD

Baris ASIKGIL, Aydin ERAR
Mimar Sinan Fine Arts University, Turkey
E-mail: basikgil@msgsu.edu.tr

There is an efficiency problem in parameter estimation of nonlinear regression models when errors are autocorrelated. In order to overcome the problem especially for autoregressive process some methods have been proposed. In this paper a modified two-stage least squares method has been developed to obtain more efficient parameter estimates than the method proposed by Gallant and Goebel (1976). The method is based on generalized least squares, but a modified weight matrix is defined by using different algorithms for parameter estimation in autoregressive process. The efficiency in parameter estimation is showed on a numerical example used by Gallant and Goebel (1976) and a Monte Carlo study.

NON-DIFFERENTIABLE OPTIMIZATION IN DATA ANALYSIS AND MACHINE LEARNING

Adil Bagirov
University of Ballarat, Australia
E-mail: a.bagirov@ballarat.edu.au

In this talk we discuss non-differentiable optimization problems in data mining and machine learning. In particular we discuss non-differentiable optimization approaches to the supervised, semi-supervised and unsupervised data classification problems. These approaches will be compared with other optimization approaches to data mining problems. We present different non-differentiable optimization algorithms for solving data mining problems. These algorithms include bundle-type, derivative free and stochastic optimization algorithms. We also present results of numerical experiments to compare non-differentiable optimization algorithms with other supervised and unsupervised data classification algorithms.

AN INCREMENTAL APPROACH FOR CONSTRUCTION OF A PIECEWISE LINEAR CLASSIFIER

Adil BAGIROV, Julien UGON, Dean WEBB
CIAO, School of Information Technology and Mathematical Sciences,
University of Ballarat, Australia
E-mail: a.bagirov@ballarat.edu.au

We consider the problem of finding piecewise linear boundaries between sets and apply it for solving supervised data classification problems. An algorithm for the computation of piecewise linear boundaries, consisting of two main steps, is proposed. In the first step sets are approximated by hyperboxes to find so-called “indeterminate“ regions between sets. In the second step sets are separated inside these “indeterminate“ regions by piecewise linear functions. These functions are computed incrementally starting with a linear function. Results of numerical experiments are reported. These results demonstrate that the new algorithm requires a reasonable training time and it produces consistently good test set accuracy on most data sets comparing with mainstream classifiers.

A NEW GLOBAL K-MEANS ALGORITHM FOR CLUSTERING LARGE DATA SETS

Adil BAGIROV, Julien UGON, Dean WEBB
CIAO, School of Information Technology and Mathematical Sciences,
University of Ballarat, Australia
E-mail: a.bagirov@ballarat.edu.au

The k-means algorithm and its variations are known to be fast clustering algorithms. However, they are sensitive to the choice of starting points and inefficient for solving clustering problems in large data sets. Recently, in order to resolve difficulties with the choice of starting points, incremental approaches have been

developed. The global k-means and the modified global k-means algorithms are based on such an approach. They iteratively add one cluster center at a time. Results of numerical experiments show that these algorithms considerably improve the k-means algorithm. However, these algorithms are not suitable for clustering very large data sets. In this talk we discuss a new version of the modified global k-means algorithm. We introduce an auxiliary cluster function to generate a set of starting points spanning different parts of the data set. We exploit information gathered in previous iterations of the incremental algorithm to reduce its complexity. Results of numerical experiments on 6 large data sets demonstrate that the new algorithm can be applied for clustering large data sets.

SURVIVAL FUNCTION ESTIMATION FOR TIME-DURATION OF STUDIES DATA: A PROPORTIONAL HAZARDS APPROACH

Dimitrios BAGKAVOS, Aglaia KALAMATIANOU
Accenture Marketing Sciences, Greece
E-mail: dimitrios.bagkavos@accenture.com

We employ the proportional hazards methodology to address a real world problem which involves survival analysis/length of stay type of data. We introduce the framework of study and then proceed to develop a stratified parametric model to fit the survival function of interest. The results are analyzed and we conclude by identifying similarities and differences among subpopulations of the available sample as these are defined by the available covariates as well as possible extensions of this research.

ANALYSIS OF AUTOMATED INTERBANK SETTLEMENT ALGORITHMS

Donatas BAKSYS, Leonidas SAKALAIUSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: donatas.baksys@ktu.lt

The aim of this paper is analysis of the automated interbank settlement algorithms. By perfecting the processing of settlements and/or developing algorithms for solving the gridlocks, or by applying the tools of refinancing and using reserves of requirements, one can change the efficiency of settlement systems. In this paper, the settlement algorithms of the Trans-European Automated Real-time Gross settlement Express Transfer system (TARGET2) are tested. The system of modelling, simulation, and optimization of settlements is used for testing of algorithms. The parameters of simulation were estimated according to the real data.

SIGNATURES OF SYSTEMS AND THEIR PROBABILISTIC AND STATISTICAL PROPERTIES

Narayanaswamy BALAKRISHNAN
University of McMaster University, Canada
E-mail: bala@univmail.cis.mcmaster.ca

In building reliability into a system, engineers must address a number of practical needs that will enable them to quantify and compare reliability in engineered systems. The practical, standardized, technical tool for characterizing reliability in systems is system signatures which was created in 1985 and since has developed into a powerful tool for qualifying reliability. It is used in all physical structures and stochastic systems where reliability is an important consideration. The representation of the reliability function of the lifetime of a coherent system as a mixture of the reliability function of order statistics associated with the lifetimes of its components is a very useful tool to study the ordering and the limiting behaviour of coherent systems. We obtain several representations of the reliability functions of residual lifetimes of used coherent systems under two particular conditions on the status of the components or the system in terms of the reliability functions of residual lifetimes of order statistics. New results and future directions for system signatures are also discussed.

EXACT BAYESIAN PREDICTION IN NON-GAUSSIAN MARKOV-SWITCHING MODEL

Noemie BARDEL, Francois DESBOUVRIES
Telecom SudParis, Evry, France
E-mail: francois.desbouvries@it-sudparis.eu

Jump Markov state-space systems are generalizations of the classical state-space model in which a so-called “jump” variable (which classically is modeled by an exogeneous Markov chain) is introduced to monitor the abrupt change of channel or kinematic characteristics and/or the presence (or absence) of a target in a tracking context. Such models have been extensively used in many applications, including track-before-detect radar processing techniques. However, it is well known that even in the linear and Gaussian case, exact propagation of the filtering a posteriori probability density function increases exponentially with time so one has to resort to suboptimal solutions such as particle filtering algorithms.

In this paper we consider a family of switch-independent alternative state-space models which still involve 3 processes (state x , observation y and exogeneous process r), but in which (y,r) is a Markov Chain. From a computational point of view, this key property enables exact Bayesian restoration, even if the measurement process and/or noises are non Gaussian.

ESTIMATION OF THE HURST INDEX OF A FRACTIONAL BROWNIAN MOTION WITH RANDOM OBSERVATION DEPENDING ON THE PROCESS

Jean-Marc BARDET, Pierre, R. BERTRAND, Medhi FHIMA
Universite Paris 1, France
E-mail: Pierre.Bertrand@inria.fr

Let X be a fractional Brownian motion (fBm) or a multiscale fBm. From the observation of a path of this process at random irregularly spaced times $t_k, k=0,1,\dots,n$, with $t_0=0$ and $t_n=T$, its spectral density or Hurst index can be estimated. When the observation times are random but independent of the process X , a standard method is the Alias Free Sampling (ASF) scheme, see Lii and Marsy, 1996. This method allows to estimate the spectral density at frequencies close to the sampling frequency when $T \rightarrow \infty$. Based on the study of two real examples (Heartbeat time series and Finance), one proposes a different modelling: the spectral density is estimated on finite band of frequencies which are smaller enough than the sampling frequency, the observation times are randomly and irregularly spaced but may depend on the process itself. In this framework, an appropriated wavelet based estimator provides a consistent estimation of the spectral density on finite bands of frequencies. The key result is to obtain a lemma bounding the discretization error. One supports this analysis by numerical simulations.

STOCHASTIC APPROXIMATION FOR SEMI-SUPERVISED NON-LINEAR SVMs BINARY CLASSIFICATION

Vaida BARTKUTE-NORKUNIENE
Institute of Mathematics and Informatics, Lithuania
E-mail: vaidaba@ukolegija.lt

In this paper, we consider the problem of semi-supervised non-linear binary classification by Support Vector Machines (SVMs). This problem is explored as an unconstrained and non-smooth optimization task when part of the available data is unlabelled. We apply non-smooth optimization techniques to classification problems where the objective function considered is non-convex and non-differentiable and so difficult to minimize. We explore and compare the properties of Simultaneous Perturbation Stochastic Approximation (SPSA) algorithms (SPSA with the Lipschitz Perturbation Operator, SPSA with the Uniform Perturbation Operator) and Standard Finite Difference Approximation (FDSA) for semi-supervised non-linear SVM classification. We present some numerical results obtained by running the proposed methods on several test problems.

APPLICATION OF ORDER STATISTICS TO TERMINATION OF STOCHASTIC APPROXIMATION

Vaida BARTKUTE-NORKUNIENE, Leonidas SAKALAIUSKAS
Institute of Informatics and Mathematics, Lithuania,
E-mail: sakal@ktl.mii.lt²

This paper is designed to present an application of order statistics provided by stochastic approximation to estimate the minimum value of the objective function and its confidence interval after a finite number of iterations. A stopping criterion - length of the confidence interval of the objective function extreme value is introduced, estimated according to order statistics of the function values obtained during the stochastic approximation. The efficiency of this approach is discussed, using the results of computer modelling.

ESTIMATION OF THE THREE-PARAMETER WEIBULL DISTRIBUTION WITH APPLICATIONS TO LARGE-SCALE DATA SETS

Vaida BARTKUTE-NORKUNIENE, Leonidas SAKALAIUSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: vaidaba@ukolegija.lt

In this paper, an approach which allows us to estimate the three parameters of Weibull distribution using order statistics is discussed. We construct numerical algorithm that takes advantage of improved analytical estimator of the shape parameter when parameters of location and scale are not known. The proposed estimator is asymptotically unbiased and consistent. We investigate the properties of this estimator in large-scale data sets.

MODELLING STROKE PATIENT PATHWAYS USING SURVIVAL ANALYSIS AND SIMULATION MODELLING

Maria BARTON, Sally MCCLEAN, Lalit GARG, Ken FULLERTON
School of Computing and Information Engineering, University of Ulster, UK
E-mail: m.barton@ulster.ac.uk

Stroke disease is the third leading cause of death in the UK, placing a heavy burden on society at a cost of 7 million pounds per year. Prolonged length of stay in hospital is considered to be an inefficient use of hospital resources. In this paper we present results of survival analysis that utilise length of stay and destination as outcome measures, based on data from the Belfast City Hospital. Survival probabilities were determined using Kaplan Meier survival curves and log rank tests. Multivariate Cox proportional hazards models were also fitted to identify independent predictors of length of stay including age, gender and diagnosis. Elderly patients showed a decreased hazard ratio of discharge. However, gender was not a significant hazard risk for length of stay in hospital. Those patients with a diagnosis of cerebral haemorrhage showed an increased hazard ratio and hence were most likely to have a shorter length of stay and to die in hospital. Those who were eventually discharged to a Private Nursing Home had the lowest probability of early discharge.

On the basis of these results we have created several groups, stratified by age, gender diagnosis and destination. These groups are then use to form the basis of a simulation model where each group is a patient pathway within the simulation. Various scenarios are explored with a particular focus on the potential efficiency gains if length of stay in hospital, prior to discharge to a Private Nursing Home, can be reduced.

APPLICATION OF SUPERVISED LEARNING METHODS TO ECOLOGICAL DATA ANALYSIS

Dalia BAZIUKAITE, Natalija JUSCHENKO
Klaipeda University, Lithuania
E-mail: dalia.baziukaite@ku.lt

Ecological data usually are complex, incomplete and have descriptive manner. Practical tasks solving is not limited to statistical data analysis, but requires a forecast as well. Described features of datasets narrow the list of possibly applicable prognostic methods.

This paper deals with an application of some supervised learning algorithms for analysis of complex, incomplete and descriptive dataset on example of biological invasion data.

COMPARISON OF DIFFERENT PANELS SORTING TASKS WITH HIERARCHICAL MULTIPLE FACTOR ANALYSIS

Monica BECUE-BERTAUT, Berenice COLMENARES, Sebastien LE
Universitat Politecnica de Catalunya, Barcelona, Spain
E-mail: monica.becue@upc.edu

Hierarchical Multiple Factor Analysis (HMFA; Le Dien & Pages, 2003a) is an extension of Multiple Factor Analysis that analyzes data where a hierarchical structure is defined upon the variables. HMFA has been used to compare sensory profiles provided by trained and non-trained panels (Le Dien & Pages, 2003b). In both works, only quantitative variables were considered.

We explore here HMFA potentialities to compare free-sorting tasks results in wine-tasting hall tests with different nationality panels. We adopt the point of view presented in (Le, Cadoret, Pages, 2008), that is, we assimilate the partition defined by each panellist to a categorical variable presenting as many categories as clusters. We consider two levels in HMFA. The different partitions constitute the sets taken into account at the first level; they are gathered by nationality to make up the second level sets. Among other graphical representations, HMFA will display a synthetic visualisation of the proximities between all the tasters, as well as between the nationalities.

HMFA is also able to jointly process free-sorting task and liking scores. At the first level, we associate two sets to each nationality, the liking-scores and the partition sets, that is, mixed data with quantitative and categorical sets. At the second level, for every nationality, both sets are gathered. The relationship between the two types of assessments within every panel as well as the comparison of these relationships through out all the panels can thus be performed.

The data refer to 8 Catalan wines, tasted in two different hall tests by, respectively, a Catalan and a French panel following a same protocol. The data analysis shows a great coincidence between the structures defined upon the wines by every panel as well as a close relationship between liking-scores and free sorting-task clusters within every panel.

1. Le Dien, S. & Pages, J. (2003) Analyse Factorielle Multiple Hiérarchique. *Revue de Statistique Appliquée*. LI (2), 47-73.
2. Le Dien, S. & Pages, J. (2003) Hierarchical Multiple Factor Analysis: application to the comparison of sensory profiles. *Food Quality and Preference*. 14 (5-6), 397-403.
3. S. Le, Cadoret M., Pages J. (2008) A novel Factorial Approach for analyzing Sorting Task data. First joint meeting of the SFC and the CLADAG. Caserta (Italie), 11-13 juin 2008.

A MONTE CARLO SEMI-MARKOV CLAIM RESERVE MODEL

Elena BIFFI, Jacques JANSSEN, Raimondo MANCA
University G. D'Annunzio of Chieti-Pescara, Italy
E-mail: raimondo.manca@uniroma1.it

In a previous paper (Biffi et al (2007)) a model useful for the claim reserve calculation was presented. This model reconstructs the probability functions by means of a Monte Carlo simulation method. The simulation procedure is applied on the evolution equations that are obtained solving a homogeneous semi-Markov process. The non-homogeneous version of the model was applied in a credit risk environment in Biffi et al (2008a) and (2008b). In this paper the claim reserve model will be generalized. Indeed an initial backward process will be attached to the homogeneous semi-Markov process. The evolution equation solution of this initial backward semi-Markov process will give the basis probabilities for the application of the Monte Carlo simulation model.

The model will be applied for the reconstruction of the probability function of a claim reserve of an insurance company. The introduction of the initial backward recurrence time process will give the possibility to consider the IBNR (incurred but not reported) claim request.

An application of the model to real data of an Italian insurance company will be presented.

1. Biffi E., Janssen J., Manca R. (2007). Un modello Monte Carlo semi-Markoviano utile alla misura della riserva sinistri. (with) Proc. VIII congresso degli attuari. Trieste.
2. Biffi E., D'Amico G., Di Biase G., Janssen J., Manca R., Silvestrov D. (2008). Monte Carlo semi-Markov methods for credit risk migration models and Basel II rules I. *Journal of Numerical and Applied Mathematics*, 96, 28-58.
3. Biffi E., D'Amico G., Di Biase G., Janssen J., Manca R., Silvestrov D. (2008). Monte Carlo semi-Markov methods for credit risk migration models and Basel II rules II. (with). *Journal of Numerical and Applied Mathematics*, 96, 59-86, 2008.

NEGATIVE DEPENDENCE AND THE SIMES INEQUALITY

Henry W. BLOCK, Thomas H. SAVITS and Jie WANG
University of Missouri, USA
E-mail: hwb@pitt.edu

It is shown that the Simes Inequality is reversed for a broad class of negatively dependent distributions. This resolves a conjecture of Sarkar (*Ann. Stat.* 26, 1998, 494-504).

ASYMPTOTICALLY ROBUST INVARIANT ALGORITHMS FOR OVERCOMING NONPARAMETRIC UNCERTAINTY OF PDF AND PARAMETRIC UNCERTAINTY OF SCALING AND SHIFT PARAMETERS

Veniamin BOGDANOVICH, Aleksey VOSTRETISOV
Novosibirsk State Technical University, Russia
E-mail: vostretsov@adm.nstu.ru

Using the principles of invariance and minimax, and Bayesian approach the method of synthesis of asymptotically robust invariant (ARI) algorithms for detection and recognition of signals with unknown parameters on the background of independent noise and signal-like interference with uncertain parameters is developed. According to this method the ARI algorithm is sought in the class of algorithms of correlation type with the threshold depending on the observed sampling. Introduction of such class permits to use the developed method of synthesis for any model of PDF with finite Fisher's information about shift. For synthesis of a robust algorithm using this class no requirements are imposed on energy parameter of a signal, it is different from the known class of algorithms of M-type.

The conditions ensuring synthesis of ARI algorithms are formulated. These conditions are fulfilled for majority of known models for nonparametric a priori uncertainty including the model with finite variance and the q-point model.

Quality of ARI algorithms is estimated by asymptotic average losses caused by omission of the signal and erroneous estimation of its parameters for a fixed level of asymptotic probability of false alarm. Owing to optimality according to minimax criterion ARI algorithms have the minimal upper boundary for average losses for any PDF of noise and for any a priori distribution of signal parameters. Owing to invariance of the algorithm to scaling and shift transformations the average losses do not depend on the parameters of signal-like interference and for a fixed PDF of noise are determined by the only parameter γ - (energy parameter of signal)-to-(scaling parameter of noise's PDF) ratio. The asymptotic probability of false alarm has a constant value for any PDF of noise, arbitrary scaling parameters and arbitrary parameters of signal-like interference.

ARI algorithms are structurally independent of all a priori unknown characteristics of signals and interference, and preserve the optimality according to minimax criterion for any distribution of signal parameters.

The upper boundary for asymptotic average losses is a function of the only argument γ and can be calculated in advance.

The obtained ARI algorithms can be used for development of various information systems where signal-noise environment is characterized by substantial a priori uncertainty.

MULTIBLOCK REDUNDANCY ANALYSIS. APPLICATION TO EPIDEMIOLOGICAL SURVEYS

Stephanie BOUGEARD, El Mostafa QANNARI
AFSSA, France
E-mail: s.bougeard@afssa.fr

For the purpose of exploring and modelling the relationships between a data table Y and several tables (X_1, \dots, X_K) measured on the same individuals, multiblock Partial Least Squares (PLS) is a regression technique which is widely used, particularly in chemometrics, sensometrics and marketing studies. It is designed as an extension of PLS regression which aims at linking two datasets. Redundancy Analysis is another popular method for linking two datasets, which is more oriented towards the Y explanation. We propose an extension of Redundancy Analysis to the multiblock setting. This method is computed by maximizing a criterion that reflects the objectives and solutions are directly derived from the eigenanalysis of a matrix. We also outline how the method is related to other multiblock methods. A main difference of multiblock Redundancy Analysis with other techniques lies in the fact that it focus on the global components rather than the partial components. Therefore, it gives valuable tools both for the prediction and the investigation of the relationships among datasets. We demonstrate the benefits of our approach on the basis of a real dataset pertaining to the field of veterinary epidemiology. The description of the links between variables, variable blocks and individuals is given through graphical displays. On the basis of a cross-validation study, we show that our approach leads to significant improvements comparing to other techniques.

CLUSTERING IN FISHER DISCRIMINATIVE SUBSPACES

Charles BOUVEYRON, Camille BRUNET
University Paris 1, France
E-mail: charles.bouveyron@univ-paris1.fr

Clustering in high-dimensional spaces is nowadays a recurrent problem in many scientific domains but remains a difficult problem. This is mainly due to the fact that high-dimensional data usually live in low-dimensional subspaces hidden in the original space. This paper presents a model-based clustering approach which models the data in a discriminative subspace with an intrinsic dimension lower than the dimension of the original space. An estimation algorithm, called Fisher-EM algorithm, is proposed for estimating both the mixture parameters and the discriminative subspace. Experiments show that the proposed approach outperforms existing clustering methods and provides a useful representation of the high-dimensional data.

FUNCTIONAL ESTIMATION OF THE RANDOM RATE OF A COX PROCESS

Paula R. BOUZAS, Mariano J. VALDERRAMA, Ana M. AGUILERA,
Nuria RUIZ-FUENTES
University of Granada, Granada, Spain
E-mail: paula@ugr.es

The rate of a Cox process is also a stochastic process that can be derived by q.m. differentiation of its random mean function. So, from a set of sample paths of the mean process we propose a numerical method to estimate the intensity preserving its monotone character on the basis of the functional PCA. A validation of the estimation method is presented by means of a simulation as well as a comparison with an alternative estimation method.

ARTIFICIAL IMMUNE RECOGNITION SYSTEMS (AIRS) FOR THE PREDICTION OF STOCK MARKET RETURN: THE CASE OF THE ISTANBUL STOCK EXCHANGE

Melek Acar BOYACIOGLU, Kemal POLAT, Seral LEEN
Selcuk University, Turkey
E-mail: melekacar@yahoo.com

Stock market prediction is important and of great interest because successful prediction of stock prices may promise attractive benefits. These tasks are highly complicated and very difficult. In this paper, we investigate the predictability of stock market return with Artificial Immune Recognition Systems (AIRS). The objective of this study is to determine whether an AIRS algorithm is capable of accurately predicting stock market return. We attempt to model and predict the return on stock price index of the Istanbul Stock Exchange (ISE) with AIRS. We use six macroeconomic variables as input variables. The experimental results show that AIRS provides a promising alternative to stock market prediction. AIRS is a useful tool for economists and practitioners dealing with the forecasting stock price index return.

MISSING VALUES IN CATEGORIZATION

Marine CADORET, Sebastien LE, Jerome PAGES
Agrocampus Ouest, France
E-mail: marine.cadoret@agrocampus-ouest.fr

Categorization is a task in which subjects are asked to group a set of object according to their similarities. This method is increasingly used but may be difficult when the number of items to categorize is high: this appears particularly true in sensory analysis when the five senses of the subjects are mobilized.

The aim of this communication is twofold: first, to propose a solution to this difficulty which consists in presenting only a subset of objects to each subject, de facto data will contain missing values; second to present an approach based on multiple correspondence analysis (MCA) to analyze categorization data with missing values. We will show how to balance according to an experimental design the absence of each object and each pair of objects for a given subset to be presented. We will also show how to take into account missing values in MCA: they can be taken into account by adding a new group to each subject, the group of the objects that haven't been presented. We will illustrate our method with real data in two ways: on the one hand, missing values will be simulated from a complete data set of 12 luxury perfumes; on the other hand, missing values will be obtained by presenting subsets of perfumes out of the 12 fragrances. The results obtained with and without missing values will be compared with multiple factor analysis (MFA).

BAYRAM: AN ARCHITECTURE FOR BAYESIAN ASSESSMENT OF RELIABILITY, AVAILABILITY AND MAINTAINABILITY OF HARDWARE/SOFTWARE SYSTEMS

Javier CANO, David Rios INSUA
University of Rex Juan Carlos, Spain
E-mail: javier.cano@urjc.es

The importance of HW/SW systems has been constantly increasing during the last decades, being present in many fields of human activity. Therefore, the assessment of reliability, availability and maintainability (RAM) of such systems has become a major topic of research, both in the industry and in the academic world, specially for safety critical systems in areas such as finance, aerospace and energy, among others. In this regard, there are several software packages available to support reasonably complex RAM analyses, although they have several limitations, being possibly the most important, the fact that they provide little support to Bayesian analysis. Being aware of such drawbacks, we are currently developing BayRAM, a JAVA package to support Bayesian RAM analysis of complex HW/SW systems, based on three key ideas: complex systems are described through series and parallel RBDs. Software block reliability modeling is based on model selection strategy over mixtures of Software Reliability Growth models, whereas hardware block RAM modeling is based on Continuous Time Markov Chains and the use of phase-type distributions.

The package includes a highly interactive graphical interface, which allows the user to intuitively design the system, and also comprises input data routines, some exploratory data analysis routines, prior specification modules and modules to undertake inference and prediction tasks. As an illustrative example, we provide the analysis of a schematic version of our university resource planner, showing the basic features and the flexibility of our package.

STATISTICAL ANALYSIS OF PROGRESSIVELY HYBRID CENSORING SCHEME AND ADAPTIVE PROGRESSIVELY HYBRID CENSORING SCHEME FROM EXTREME VALUE DISTRIBUTION

P.S. CHAN, H.K. NG, C.T. LIN, M.Y. MAK
Chinese University of Hong Kong
E-mail: s0702016@sta.cuhk.edu.hk

In life-testing or reliability experiments, progressively hybrid censoring schemes are commonly used to retain more samples from the experiment. It is because they allow for the removal of samples at points during the experiment. In this thesis, the Progressively Hybrid Censoring (PHC) influential method and the Adaptive Progressively Hybrid Censoring (APHC) influential method from extreme value distribution are studied. For the PHC scheme, the number of effective sample size m , the experimental time T and the progressively censoring scheme (R_1, R_2, \dots, R_m) are fixed in advance. The experiment terminates either at a pre-specified time T or at the time when the m^{th} failure occurred. For the APHC scheme, the values of R_i may change during the experiment. If the experimental time exceeds the prefixed time T but the number of observed failures does not reach m , we would want to terminate the experiment as soon as possible by adjusting the R_i 's.

Based on extreme value distribution model, we derive the maximum likelihood estimators and the approximate maximum likelihood estimators of the unknown parameters. The biases and mean square errors of the parameter estimates are studied through Monte Carlo simulation. Various interval construction methods such as asymptotic confidence intervals, percentile bootstrap confidence intervals and bootstrap-t confidence intervals are proposed. Then these methods are compared with their coverage probabilities and expected widths using simulation. The expected total time of these two progressively hybrid censoring schemes (R_1, R_2, \dots, R_m) are also worked out.

A TWO-DIMENSIONAL ANALYTIC SOLUTION FOR ROTATION IN MULTIPLE CORRESPONDENCE ANALYSIS

Marie CHAVENT, Vanessa KUENTZ, Jerome SARACCO
University of Bordeaux, France
E-mail: vanessa.kuentz@math.u-bordeaux1.fr

Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) are well-known multivariate methods for statistical description of respectively quantitative and qualitative data. A rotation may then be applied to the principal component scores matrix. The definition of a criterion enables to achieve simple structure, thereby simplifying the interpretation of the results. A two-dimensional analytic solution has been proposed for varimax rotation in PCA. We propose here a two-dimensional analytic solution for rotation in MCA using a varimax-based criterion relying on correlation ratio. An application on a real data set shows the possible benefits of using planar rotation in MCA.

ADVANCED SYSTEM FOR AUTOMATED INTERVENTION IN RISK AND HAZARDOUS SITUATIONS

Oana CHENARU, Luiza OCHEANA, Gheorghe FLOREA
SIS S.A., Luiza, Gheorghe, Romania
E-mail: oana.chenaru@sis.ro

Most of industrial environments share the need for a higher level decision and control system to improve and optimize plant operation. Different research projects have been developed worldwide in this area and their results will be considered in designing and implementing the system.

Based on the actual trends in the industrial environment of using models to describe the dynamic behavior of the process, we want to develop an automated intelligent system able to predict a potential risk and capable of interfering in the process functioning in a way that will minimize the potential hazards. The system will use real-time process data that will be analyzed, the process model will be adjusted and the possible risks will be identified. The implementation of such a system will be discussed, as well as the issues that arise when a remote computer has the control over a plant.

CAN WE UTILISE THE NEURAL-FUZZY SYSTEMS IN THE ARTIFICIAL FIELDS?

Lidia CRISTEA, Enona CRISTEA
Politehnical University of Bucharest, Romania
E-mail: lidiacristea@zappmobile.ro, ugc_rector@ugc.ro

The communique of the Conference of European Ministers Responsible for Higher Education from 19-20 May 2005 [6] states in the first chapter that we must introduce "the innovative teaching and learning process that Europe needs". Also the quality assurance is made by "a high degree of cooperation and networking." The main actor, the student must be involved in the internal mechanisms of quality assurance and maybe the artificial learning can give the possibility to externalize the quality assurance. The new system of lifelong learning in higher education gives the best qualification for the European market labor.

The Bergen Conference states the principles of quality and transparency and the duty to cherish the heritage and cultural diversity in the use of a knowledge-based society.

The present paper tries to express the possibility to use the hybrid systems neural-fuzzy to exemplify the new concept of artificial learning in the context of the new Higher Space of Education [2]. We utilize the Soft Computing notion, introduced by Lotfi Zadeh [3] with the target to explore the tolerance between the hazard and the truth, for a good flexibility, low cost and good links with the reality. The Soft Computing has a good partnership between the neural networks, genetically algorithms, fuzzy logical and probabilistic rationaments.

In this technique the basic is the human thinking, all this relations are in synergism, each partner has his own contribution with the own advantages and techniques for the best solutions. The neural networks have the capacity to learn and to improve a great adaptability, the fuzzy logical offers the approximation

possibility, in the same time the genetically algorithms can realize systematical research of the optimal solution. We made a comparison between the fuzzy systems and the neural networks with the advantages and disadvantages in the learning process [4,5]. For example the neuron networks the learning process is slowly and we can consider the system like a black box only with the input and output. If we put together the both systems we have more advantages and less disadvantages. The hybrid systems study is the neural network fuzzy, the fuzzy perceptron, the fuzzy neural systems, the hybrid systems neural-fuzzy and the adaptive network based fuzzy interference system (ANFIS).

The fuzzy systems and the neural networks are the numerical estimators and in the same time dynamics systems which present the capacity of development the brains of the potential students in the uncertain area of higher European education. That is the target of the soft-computing, which try to simulated the human performance brain mixing the new techniques like neural networks, fuzzy systems and genetically alghoritms. The fuzzy systems are good when we are base on the expert systems and the neural networks are good if exist a lot of data based .The fuzzy systems give a symbolic -qualitative mechanism and the neural networks give a numerical-quantitative mechanism.

References

- 1.The European Higher Education Area-Achieving the Goals. –Communique of the Conference of European Ministers Responsible for Higher Education, Bergen, 19-20 May 2005.
- 2.Lidia Cristea ,E.Jagodnik, The continuos learning formation based on the Fuzzy systems,UNISO 2007, Versailles, 9-12 July, 2007,France.
- 3.L.A.Zadeh:Fuzzy Sets, Information and Control,8:338-353,1965
- 4.A.Abraham:Beyond Neuro-Fuzzy Systems.Rewiews, Prospects,Perspectives and directions,Monash University, Australia,wwwbs.monash.edu.au/staff/Abraham/mendel2001.ppt.
- 4.B.Qiu:Sugeno Type Fuzzy Interference, Monash University,Australia ,www.csse.monash.edu.au/~bq/teaching/cse5301/ch8.ppt.

MODELLING LITHUANIA'S INFLATION: APPLICATION OF VAR MODEL

Ana CUVAK, Zilvinas KALINAUSKAS
Vilnius Gediminas Technical University, Lithuania
E-mail: Ana.Cuvak@fm.vgtu.lt

Inflation is one of the crucial modern macroeconomic problems. In our days inflation theme is very relevant. Time series models, linear regression models and vector autoregression model (VAR) can be used to model and forecast inflation processes. This paper examines the Lithuanian consumer price inflation using a modern stationary time series and econometric theory. The vector autoregression model is proposed for inflation modeling. Theoretical aspects of model estimation are reviewed: time series stationarity, model identification, parameter estimation, model usage and forecasts. The stationarity of HPCI index and exogenous variables are analyzed using the Augmented Dickey - Fuller (ADF) test. Vector autoregression model of Lithuaninan inflation processes is investigated and proposed for inflation modeling. Obtained model is used for forecasting purposes and showed fairly high accuracy of inflation forecast in the nearest 12 month period.

QAMML DISTRIBUTIONS : A SPELL TO BREAK THE CURSE OF DIMENSIONALITY FOR STOCHASTIC PROCESS

Etienne CUVELIER, Monique NOIRHOMME-FRAITURE
FUNDP, Belgium
E-mail: ecu@info.fundp.ac.be

Define and calculate a probability distribution for a stochastic process, with domain on a real interval, is very difficult without any assumption. The Brownian motion is a very well known assumption which permits to define more easily a probability distribution. If no assumption is made, the only way to deal with this problem is to use the Kolmogorov's extension theorem. This theorem is based on the classical property which links a joint distribution to its marginal distributions. The practical consequence of this result is the

usual use of the finite dimensional distributions to characterize a stochastic process. Thus, working with a finite dimensional distribution of dimension N can be seen as working with an approximation of the N th degree. The curse of dimensionality limits the precision of a such approximation.

In this paper, instead of using a classical joint distribution directly linked to its margins, we define a new kind of joint distribution, which conserve the stochastic quantiles. A solution is given using a quasi arithmetic mean with the generator function of an Archimedean copulas : the QAMM distributions. This new kind of joint distribution leads to a new solution to define a probability distribution for stochastic process : the QAMML distributions. These distributions are directly defined in the infinite dimensional space in which the stochastic processes lie.

ESTIMATION OF CHOSEN POPULATION PARAMETERS

Wieslawa DABALA
Public Opinion Research Centre, Poland
E-mail: w.dabala@cbos.pl

Population was divided into strata and two-stage sampling scheme was used in sample selection. Primary sampling units (psu's) were chosen with application Sunter's scheme. Secondary sampling units (ssu's) were selected from previously sampled psu's using simple sampling without replacement. The response rates in strata were differenet. As estimators of the population parameters, statistics adjusted by weights were used. The calculation of weights included: probabilities of sampling units, response rates in strata and external data for correcting proportions. Chosen estimated population parameters were: fraction, arithmetical mean, median. The estimation of variance of estimators was carried out with simplified methods of resampling.

HURST ANALYSIS OF BALTIC SECTOR INDICES

Svetlana DANILENKO
Vilnius Gediminas Technical University, Lithuania
E-mail: sdanilenko@gmail.com

The Hurst exponent is a statistical measure used to classify time series. Using the Hurst parameter processes are classified into long-term dependence, antipersistence and white noise processes. R/S analysis method is one of the many methods that evaluate the Hurst exponent. This method uses the rescaled range statistic (R/S statistic). The literature that focuses on particular sector indices is relatively limited, therefore valuating whether long-term dependence exists in specific sectors of the economy is particularly useful. Investigation object - the Baltic sector indices. The latter represent tendencies of different sector activity indices in the stock market. The work concentrates on calculating the Hurst exponent, evaluated Hurst exponents of the Baltic sector indices are given for different periods of time.

ON STOCHASTIC MULTICRITERIA DECISION ANALYTICS AND ARTIFICIAL INTELLIGENCE FOR EFFICIENT STOCK TRADING

Gordon DASH, Nina KAJIJI, John FORMAN
University of Rhode Island, USA
E-mail: ghdash@uri.edu

Stock price modeling has taken on increasing importance as the complexities of global markets continue to evolve. These complexities have encouraged the ongoing development of mathematical modeling of stock price behavior. Stochastic methods have become increasingly important as they rely upon random or pseudorandom methods to determine an asset's future price. This paper joins stochastic multicriteria decision analytics with neural-network based modeling to present a mathematical model to predict the expected probability of profitable trading for individual securities. The policy inference produced by this research is directed at active market traders; or, more specifically, to the users of automated trading systems. An automated trading system is a programmed system that operates independently for the purpose of

generating specific trade recommendations for any grouped security. The paper proceeds in phases. To this end, the initial phase introduces the Dash-Kajiji-4 Autotrader system (DK4AT). The DK4AT is a recently released dual neural network framework stock price (return) forecasting system that combines end-of-day (EOD) and intra-day forecasts to enumerate transaction decisions (sell, buy, hold) for both long- and short-portfolio positions. The DK4AT has the ability to incorporate any number of advanced trading rules. All actual trading results for 930 securities, 2 market indices over a 1-year daily trading period were generated by the artificial intelligence of the DK4AT. The next phase of the paper relies upon MCDA theory to classify the DK4AT generated trades into ten unique clusters. This is accomplished by use of a Kohonen self organizing map (K-SOM) and firm fundamental characteristics as referenced in extant literature. Upon generating K-SOM scores, the individual firm related scores are transformed to posterior probability estimates by invoking the softmax activation function. For any given security, the result is to produce the posterior probability of profitable trading within a decile classification level. Lastly, in the final phase of the paper we seek to extrapolate these findings into a forecast model. That is, we sort (estimate) the expected probability of future stock trading classification by a employing a Cobb-Douglas (CD) production model. The parameters of the CD model are extracted by applying the Kajiji-4 radial basis function (RBF) artificial neural network (ANN). In a sub-note, the research argues for equivalence between RBF ANN output weights and estimated production elasticity estimates. As such, the weights measure the percentage change in the target variable (the softmax probability used to classify individual stocks into decile groups) given a 1-percent change in a predictor variable (e.g., fundamental variable such as EPS), ceterus paribus. The contribution of this research is directly applicable to any automated trading system. Using the stochastic MCDA framework presented in this research, stock trading decision-makers who implement automated trading systems would be able to evolve a predictive model to estimate the expected profitability ranking for any traded security. Over time, the obtained result would be to produce greater trading returns with a decreasing risk component.

TESTING FOR SPECIFIC MODELS IN REGRESSION ON FUNCTIONAL VARIABLE

Laurent DELSOL, Frederic FERRATY, Philippe VIEU
Institut de Statistique, University Catholique de Louvain-la-Neuve, Belgium
E-mail: laurent.delsol@uclouvain.be

Many practical problems involve functional type phenomena (growth curves, spectrometric curves, sound records, handwritten letters, images...). To take into account the regularity of this kind of data, it is often relevant to propose statistical tools adapted to models that involve functional random variables. Functional statistics is an interesting topic of modern statistics for both practitioners and theorists. We consider in this talk the issue of predicting a real value of interest from the observation of a functional data. Before any estimation of the link between these two quantities, it seems relevant to wonder if this link exists and if it has a specific structure. In this talk we discuss a new way to construct structural tests to answer to these questions. We present a general theoretical result and discuss how it can be used to make tests for no effect, linearity, single index model. We consider nonparametric alternatives and construct our test statistics from a functional kernel estimation of the difference between the true regression operator and an estimator specific to the sub model we want to test. Some bootstrap procedures are proposed to compute the threshold. Finally, we observe from some applications that getting additional information on the structural nature of the link we want to estimate allow to justify or reject some specific estimators and often leads to better prediction results.

EQUITY RISK FOR INSURANCE LIABILITIES : THE DURATION SYNDROM

Pierre Devolder
Universite Catholique de Louvain, Belgium
E-mail: Pierre.Devolder@uclouvain.be

Insurance liabilities have a strong investment component and financial risk is crucial especially in life insurance as proved for instance in the Solvency 2 first market results. We consider here the equity risk

induced by investment in stocks. Present rules proposed in Solvency 2 penalize strongly for an insurer this kind of investment. On the other hand, it is well known that stocks produce generally better returns on a long-term horizon. Classical risk measurement tools, such as value at risk, can be used in order to assess the “stochasticity“ generated by these investments. Another risk measure more often used in insurance and based on the concept of probability of ruin can also be proposed to analyze these risks. The purpose of this presentation is to compare these two approaches in a Geometric Brownian environment. We stress in particular the importance of the duration of the insurance liabilities in order to measure the real equity risk, in comparison with the present proposed rules of Solvency 2 based only on a one-year horizon. This duration aspect is of first importance especially in pension business where the long term is trivially the rule!

COMPARING STATISTICAL SIMILARITY MEASURES IN STYLISTIC DOCUMENT CLUSTERING

Liviu DINU, Marius POPESCU
University of Bucharest, Romania
E-mail: popescunmarius@gmail.com

Computational stylistics investigates texts from the standpoint of individual style (author identification) or functional style (genres, registers). Because in all computational stylistic studies / approaches, a process of comparison of two or more texts is involved, in a way or another, there was always a need for a distance function to measure similarity (more precisely dissimilarity) of texts from the stylistic point of view. The goal of this paper is to compare a set of distance / similarity measures, some motivated statistically others motivated stylistically, regarding their ability to reflect stylistic similarity between texts.

As style markers we used the function words frequencies. Function words are generally considered good indicators of style because their use is very unlikely to be under the conscious control of the author and because of their psychological and cognitive role. The distance / similarity between two text will be measure as distance / similarity between the function words frequencies corresponding to the respective texts. For this study we selected some similarity / distance measures. We started with the most natural measures: euclidean distance and (taking into account the statistical nature of data) Pearson's correlation coefficient. Because function words frequencies can be viewed as ordinal variables we also considered for comparison some specifically similarity measure: Spearman rank-order coefficient, Spearman's footrule, Goodman and Kruskal's gamma, Kendall's tau. Finally, we added a stylistically motivated similarity measure: Burrows's delta, that has interesting statistically interpretations. To assess the ability of these distance / similarity functions to capture stylistic similarity between texts, we plugged them into a standard hierarchical clustering algorithm and test it on a collection of 21 nineteenth century English books. The family trees thus obtained can reveal a lot about the distance / similarity measures behavior.

A LANGUAGE INDEPENDENT KERNEL METHOD FOR CLASSIFYING TEXTS WITH DISPUTED PATERNITY

Liviu DINU, Marius POPESCU
University of Bucharest, Romania
E-mail: liviu.p.dinu@gmail.com

Kernel methods prove to be very effective in many computational linguistics and text analysis tasks. Authorship identification is not an exception. The success of kernel methods in this domains is due to the fact that they are state of the art techniques in machine learning and also because they allow the use of specialized kernels that can directly manipulate the text (without explicitly embedding the text in a numerical feature space) and can incorporate prior knowledge about the problem. The main goal of this paper is to investigate the behavior of string kernels in author identification for texts with disputed paternity. First, we tested the method on the well known case of authorship of disputed Federalist papers. Doing this, not only that we will assess the performance of string kernels on a data set that since the work of Mosteller and Wallace became a sort of benchmark in authorship identification, but also we can compare our results with results obtained by other researchers with other methods on the same data set, for example: results achieved by other kernel methods or by Markov chains based methods.

We also tested the method on an example from Romanian literature: did Radu Albala found the continuation of Mateiu Caragiale's novel "Sub pecetea tainei", or did he write himself the respective continuation?

A second objective of the paper is to see if the performance achieved is inherent to the kernel type, that is if the performance remains the same if the kernel is used in conjunction with different kernel methods. We tested it with Support Vector Machines (SVM) and Kernel Fisher Discriminant(KFD). We will also discuss why we consider that string kernels are adequate to authorship identification and interpret the performance of string kernels vis-a-vis of the performance of Markov chains based methods in the light of the relation that exists between string kernels and Markov models via the so called Fisher Kernels, kernels derived from probabilistic generative models.

EXTENDING THE SCOPE OF SELF-CORRECTING

Shlomi DOLEV and Sergei FRENKEL
IPIRAN, Moskow, Russia
E-mail: fsergei@mail.ru

Random self-reducibility to obtain self-correcting in a probabilistic way has been used in the scope of finite fields. This research presents ways to extend the scope of self-correcting to other functions, such as functions over the real numbers. The extension is based on segmentation, approximation by polynomials, error correcting codes, and interpolation. In particular, we suggest to use a known reconstruction of real numerical function. The applications of the new technique include coping with hard-errors, design-errors, manufacturing errors in hardware. In addition, sanity checks and corrections of the output of analog devices.

RANDOM SERIES WITH TIME-VARYING DISCOUNTING. THE SYMMETRICAL CASE

Doncho DONCHEV
Sofia University, Bulgaria
E-mail: doncho@fmi.uni-sofia.bg

Let ξ_1, ξ_2, \dots , be i.i.d. random variables. Consider the series $1 + \sum_{n=1}^{\infty} \prod_{i=1}^n \xi_i$. It turns out that its partial sums are equal to some random determinants. This allows us to represent the original series as an infinite product of random variables that form a Markov chain. The limit behaviour of the series depends on the stationary distribution of this chain. In our recent paper we have studied the relationship between this distribution and the distribution of a non-negative ξ_i . Here, we are extending our approach to symmetrically distributed ξ_i .

IDENTIFYING TRENDS AND OUTLIERS IN NYSE INDUSTRIAL CATEGORIES

Andreea DRAGUT, Maria DRAGUT
University Aix-Marseille II, Maria, France
E-mail: andreea.dragut@univmed.fr

A normal consequence of the stock market behavior is the increased difficulty in attracting and retaining investors. In 2008 survey done for the NYSE EURONEXT 2009 CEO Report: Managing During Economic Turbulence, nearly one out of three business leaders indicates that it has become more difficult to attract investors, compared with one in five who pointed out this issue in 2007. Instead of entering a position in the hope of being at the onset of a new trend and then having the patience to hold the position long enough to allow the profit to accumulate into a large winner, most market participants are shaken out on a short-term. Thus the companies depend more on the perception of their stock on the market.

Every stock sold on the New York Stock Exchange is classified into an industrial category by NYSE itself. A company is identified with an industry based on its primary activities. With respect to the short term investing activity, the stock buyers are nowadays largely influenced by the risk exposure of certain branches of a company activity, not necessarily by the area from which that company derives the largest share of its revenue.

Considering a stream of historical price data per each traded stock, we propose a linear time clustering algorithm which investigates to what extent stocks in a certain industry move as a group. The market tendency to move as a whole is accentuated in financial crisis times. Thus, to see the differences in groups, the overall market movement can be partially removed using a classical time-based normalization. More computationally effective, this approach allows as well for an immediate reaction to the changes in the data stream based on subtler trends. To assess the market trend in a certain time window, one can look at how many transactions across all stocks occur on upticks versus downticks as a way of assessing whether buyers or sellers are more motivated. This statistic is called the NYSE tick. It is calculated by the exchange 10 times per minute (every six seconds). However, a number of issues need to be investigated since the NYSE tick, as a distribution, used not have a perfect zero mean. There was a positive bias to the series. That bias has since been reduced by the elimination of the uptick rule for short-selling, which might be reinforced once more.

THE INFLUENCE OF THE ANISOTROPY RATIO ON THE EXPECTED ERROR RATES

Lina DREIZIENE, Kestutis DUCINSKAS
Klaipeda University, Lithuania
E-mail: kestutis.ducinkas@ku.lt

The error rates in supervised classification of Gaussian random field observation into one of two populations specified by different means and common stationary geometrically anisotropic covariance are considered. The new approximation of the expected error rate for the plug-in Bayes discriminant function based on ML estimators of unknown parameters is proposed. The influence of the anisotropy ratio on the accuracy of the proposed approximation is evaluated numerically for small training samples and exponential structure of spatial correlation.

ERROR RATES IN CLASSIFICATION OF OBSERVATION OF MULTIVARIATE GAUSSIAN RANDOM FIELD WITH FACTORIZED COVARIANCE

Kestutis DUCINSKAS
Klaipeda University, Lithuania
E-mail: kestutis.ducinkas@ku.lt

The problem of discrimination between two different populations of multivariate Gaussian random fields, specified by different regression mean models and by factorized common covariance function. The goal is to classify the new observation into one of two prescribed populations by using spatially dependent training sample. Unknown means and common variance-covariance matrix as are estimated by ML method from training sample. These estimators are plugged in the Bayes discriminant function. The asymptotic approximation of the expected error rate associated with plug - in Bayes discriminant function is derived. Numerical analysis of the accuracy of the proposed approximation in the finite training sample size case is carried out.

OPTIMAL SAMPLING DESIGN FOR THE GLS ESTIMATORS OF THE MEAN PARAMETERS FOR THE STATIONARY SPATIAL RANDOM PROCESS

Kestutis DUCINSKAS, Ingrida BORISENKO
Klaipeda University, Lithuania
E-mail: ingrida@ik.ku.lt

The performance of estimators of the mean parameters for the stationary spatial random process (SRP) are effected substantially by the spatial configuration of locations where observations are taken (sampling design). Criteria functions for D-optimal and A-optimal sampling designs for GLS estimators of mean parameters based on spatial semivariogram are derived. In the case of stationary SRP with exponential semivariogram the second-order neighborhood scheme are numerically investigated.

THE MINIMUM DENSITY POWER DIVERGENCE APPROACH IN BUILDING ROBUST REGRESSION MODELS

Alessandra DURIO, Ennio Davide ISAIA
University of Turin, Italy
E-mail: isaia@econ.unito.it

In applied statistics regression is certainly one of the widespread tool in establishing the relationship between a set of predictors and a response variable. However, in situations involving the study of large data sets a careful data preparation is not feasible and data may hence be contaminated by a substantial number of outliers. If so, the estimates of the parameters of the regression model according to the Maximum Likelihood criterion are fairly unstable. Since outliers can play havoc with standard statistical methods, many robust estimators have been proposed since 1960 to be less sensitive to outliers. In this paper we investigate the use of the Minimum Density Power Divergence criterion as a valuable tool for useful parametric regression models building. More precisely, we introduce and discuss an intuitive procedure which relies on an index of similarity between estimated regression models and on a Monte Carlo Significance test of hypothesis. The procedure we suggest allows (a) to verify the presence of outliers in the data and, if they are present, (b) to choose the best tuning constant for the Minimum Density Power Divergence estimators. Theory is outlined and numerical examples, featuring several scenarios, are provided and commented.

INFLUENCE OF THE NEURON ACTIVATION FUNCTION ON THE MULTIDIMENSIONAL DATA VISUALIZATION QUALITY

Gintautas DZEMYDA, Sergejus IVANIKOVAS, Viktor MEDVEDEV
Institute of Mathematics and Informatics, Lithuania
E-mail: viktor.m@ktl.mii.lt

This paper analyzes the visualization of multidimensional data using feed-forward neural network. The dependence of the visualization quality on the configurations of the neuron activation function is presented. We investigate the unsupervised backpropagation algorithm to train a multilayer feed-forward neural network to perform the Sammon's nonlinear projection. In this paper we propose the way to speed up the network training and improve the visualization quality by adjusting the parameters of neuron activation function.

CONTINUOUS SAMPLING PLAN UNDER QUADRATICALLY VARYING ACCEPTANCE COST

Mavroudis ELEFThERIOU and Nicolas FARMAKIS
Aristotle University of Thessaloniki, Greece
E-mail: melefth@gmail.com, farmakis@math.auth.gr

The modern quality control technology is mainly based on the variations of the Continuous Sampling Plans. After much experimentation, the necessity of an economic approach of the Continuous Sampling Plans became obvious. In this paper a new idea is presented to study the behaviour of Continuous Sampling Plans in terms of economic costs. There are three kinds of economic costs: the inspection cost, the replacement cost and the acceptance cost of an item. The usual approach mentioned in the relative literature is to consider the acceptance cost constant because it is difficult to assess its behaviour. Part of the new idea of this paper lies on the fact that the acceptance cost is considered to be a quadratically varying in an attempt of the authors to interpret better complicated product lines. The new approach of studying the Continuous

Sampling Plans is implemented for reasons of simplicity on the CSP-1. Simulations are also provided to verify and illustrate the advantages of the new technique.

A NEW APPROACH FOR OUTLYING RECORDS IN BIOEQUIVALENCE TRIALS

Denis ENACHESCU, Cornelia ENACHESCU
University of Bucharest, Romania
E-mail: denaches@fmi.unibuc.ro

It is well known that outliers can have a significant influence on the conclusion of the bioavailability/bioequivalence study. In order to counter the limitations of the actual inferential methods, the present communication applies some exploratory data analysis (EDA) techniques, such as Andrews curves, Principal Component Analysis (PCA) and Projection Pursuit (PP) for detection of discordant records. Although computer intensive, the above EDA techniques detect one or more outliers in multidimensional space and does not require knowledge about data distribution.

A COMPARATIVE STUDY OF CLASSIFICATION MODELS

Birsen Eygi ERDOGAN
Marmara University, Turkey
E-mail: birsene@marmara.edu.tr

With the last financial crisis it became obvious that the risks taken by the largest banks and investment firms in the global world were so excessive and that they threatened to bring down the financial system itself. Now, once again, it is very important to evaluate the likelihood that a bank may go bankrupt. There are many complex mathematical models for measuring the risk in financial systems. These models are built around statistical ideas and probability theories, assuming that the best decisions are based on quantification and numbers, determined by the patterns of the past. This study aims to give a comparison in predicting bank performances between various classification models. Three simple (The Naive Rule, Naive Bayes, and k-Nearest Neighbors) and two multivariate classification methods (Discriminant Analysis and Logistic Regression) are being compared and their classification performances are presented. Financial ratios were used to facilitate comparisons across firms of various sizes.

DEALING WITH OVERDISPERSION IN LOGISTIC REGRESSION

Birsen EYGI ERDOGAN
Marmara University, Istanbul, Turkey
E-mail: birsene@marmara.edu.tr

Overdispersion or extra binomial variation is the presence of greater variability in a data set than would be expected based on a given simple statistical model. Almost any cause of overdispersion results when experimental units are not homogeneous. Thus, overdispersion is a very common feature in applied data analysis because in practice, populations are frequently heterogeneous contrary to the assumptions implicit within widely used simple parametric models. By analogy with linear regression or simple analysis of variance one can assume that if there is nonhomogeneity in experimental units the logistic regression analysis may lead to incorrect tests and an improper estimate of residual variance. In this study we will analyze what causes overdispersion and what is its impact on the logistic regression results for the case of bank classification.

COMPARISON OF EM AND SEM ALGORITHMS IN POISSON REGRESSION MODELS

Susana FARIA, Gilda SOROMENHO
DMCT, University of Minho, Portugal
E-mail: sfaria@mct.uminho.pt

In this paper, we propose to compare two algorithms to compute maximum likelihood estimates of the parameters of a mixture Poisson regression models: the EM algorithm and the Stochastic EM algorithm. The comparison of the two procedures was done through a simulation study of the performance of these approaches on simulated data sets. Simulation results show that the choice of the approach depends essentially on the overlap of the regression lines and the initialization of the algorithms.

DUAL MFA VERSUS A TWO STEP FACTOR METHOD: APPLICATION TO AN ON-LINE MARKET SURVEY

Karmele FERNANDEZ-AGUIRRE, M. Isabel LANDALUCE-CALVO,
Juan Ignacio MODROS-HERREN
E-mail: karmele.fernandez@ehu.es

This paper aims to compare to ways to analyze by means of MFA groups of individuals in the case where all the variables are qualitative or categorical. First we consider Dual Multiple Factor Analysis (DMFA) taking into account the equivalence between Multiple Correspondence Analysis and non-normed Principal Component Analysis (PCA) of the indicators variables weighted in a convenient way. As a second possibility we consider a two step method that successfully combines Multiple Correspondence Analysis and MFA. We think that this procedure of comparing tables of categorical variables with groups of individuals has the advantage of letting the data suggest which the most relevant factors for each table are, which their relative importance facing a global analysis and how close are among them. As it is well known, when applying MFA we can obtain relationships among cases, variables and cases, and groups. We present results obtained from a market research survey carried out at the University of the Basque Country (UPV/EHU). This institution, as a part of a large project which main aim is revamping its corporate image, is about launching a corporate shop (also considered as a gift or souvenir shop). With this objective an on-line survey is carried out at the UPV/EHU to better know its members' perceptions about the institution, and to know the acceptability and the potential success of the corporate shop. Among the variables there are 21 questions on the valuation (from 1 to 4) of the same number of products (shown in a photo).

The survey is focused on the internal agents, classified in three groups: students, academic staff and administrative and service staff. In this sense, it is necessary to point out that differencing between customers and employees at a university is by no means easy. Firstly, the relationship student-university is not the usual relationship customer-corporation. Secondly, members of the academic staff segment can be considered both as internal customers and as employees. With respect to administrative and service staff, some studies reveal the lack of integration/identification of this segment in the university organizational culture, organized around teaching and research. The characteristics of the data, arranged into several groups, make it advisable the use of some group methods us the multiple table methods cited above.

A SMOOTHING ALGORITHM FOR THE INTENSITY OF A DOUBLY STOCHASTIC POISSON PROCESS

Rosa M. FERNANDEZ-ALCALA, Jesus NAVARRO-MORENO, Juan C. RUIZ-MOLINA
Department of Statistics and Operations Research. University of Jaen, Spain
E-mail: rmfernan@ujaen.es

A recursive algorithm is obtained for the fixed-point smoothing estimator of the intensity process of a doubly stochastic Poisson process. The proposed methodology can be applied under the single assumption that the covariance function of the intensity process to be estimated is finite-dimensional. Thus, the provided solution becomes an alternative approach to the traditional estimation algorithms with more restrictive hypotheses.

MULTIPHASIC INDIVIDUAL GROWTH MODELS IN RANDOM ENVIRONMENTS

Patricia A. FILIPE, Carlos A. BRAUMANN, Carlos J. ROQUETE
Universidade de Evora, Portugal
E-mail: pasf@uevora.pt

In early work, we have studied, as general models for the evolution of an individual size in a random environment, stochastic differential equations of the form $dY(t) = b(a-Y(t))dt + s dW(t)$, where $Y(t)=h(X(t))$, h being a strictly increasing continuously differentiable function, $X(t)$ the size of an individual at time t , $W(t)$ the standard Wiener process, b the growth coefficient and $a=h(A)$, where A is the average asymptotic size. The s parameter measures the intensity of the effect of random fluctuations on growth. We have applied this monophasic model, in which there is only one functional form describing the average dynamics of the complete growth curve, and studied the estimation issues.

Here we present the generalization of the above stochastic model to the multiphasic case, in which we consider that the growth coefficient b assumes different values for different phases of the individual life. For simplicity, we consider two phases with growth coefficients b_1 and b_2 . Results and methods are illustrated using bovine growth data.

STOCHASTIC MODELS FOR RELIABILITY OF PARALEL SYSTEM WITH PHYSICAL INTERACTIONS OF ITS COMPONENTS

Jerzy FILUS, Lidia FILUS
Oakton Community College, USA
E-mail: jfilus@oakton.edu

The goal of the presentation is to describe the procedure, that in some cases enables one to construct the bi-variate probability distributions of the components life times as the system reliability models. The procedure consists of two stages. At first stage the two components e_1, e_2 are tested in separation of each other in special "laboratory ("off-system") conditions". It is assumed that, as a result of applying statistical methods, one obtains probability densities, say $f_1(x_1, c_1), f_2(x_2, c_2)$ of the component life-times T_1, T_2 , respectively. In that laboratory conditions the life-times T_1, T_2 , are stochastically independent. At the second stage the components are assumed to be installed into a real system, where some physical phenomena associated with operation of one of the components contribute to failure mechanism of the other. The so created additional stresses put on the component, say e_2 changes the component's original probability density $f_2(x_2; c_2)$ by affecting its scalar or vector parameter c_2 . An "amount" of this affection, (for example, a degree of accumulation of resulting micro-damages in e_2) depends on time of the components' interaction, that is (under simple assumptions) on a realization x_1 of the e_1 component's random life time X_1 . In such a situation we may assume, that the stochastic dependence of the life time X_2 on the other "unchanged" time $X_1 = d T_1$ can analytically be expressed as a (continuous) function $c_2(x_1)$. This approach allows to define the conditional pdf, of X_2 , given $X_1 = x_1$ that is given by the equality: $g_2(x_2 | x_1) = f_2(x_2, c_2(x_1))$. Since, according to adopted in the full text assumptions, the pdf of the marginal variable X_1 is invariant under the transition $T_1 \rightarrow X_1$, (i.e., we have the equality $f_1(x_1) = g_1(x_1)$) we obtain, as stochastic model of the system's reliability, the joint bivariate pdf $h(x_1, x_2)$ of the component life times X_1, X_2 in the form: $h(x_1, x_2) = g_2(x_2 | x_1) g_1(x_1)$.

A MULTIVARIATE LIFETIME MODEL BASED ON GENERALIZED EXPONENTIAL DISTRIBUTIONS

Manuel FRANCO, Juana-Maria VIVO
University of Murcia, Spain
E-mail: jmvivomo@um.es

In this work, we study a multivariate lifetime model based on latent random variables with generalized exponential distributions, which is an extension of the new class of bivariate distributions given by Sarhan and Balakrishnan (2007). We determine the joint survival function of the n-dimensional extension, and then the corresponding marginal survival functions. The series and parallel systems from this multivariate distribution are also derived. Moreover, we discuss the multivariate ageing classification of this model as well as the ageing of its marginals and series and parallel systems.

EXACT RESULTS FOR VARIABLE WAGER HI-LO

James FREEMAN
Manchester Business School, United Kingdom
E-mail: jim.freeman@mbs.ac.uk

HI-LO - a gaming routine commonly exploited by the gambling industry - has been analysed from a number of different perspectives in recent years, including decision analysis (DA) and stochastic dynamic programming (SDP). Ironically, although SDP is generally seen as having the advantage over DA, the SDP procedure currently in use, is only capable of delivering approximations of HI-LO gamblers' financial expectations. Now, based on a new formulation, an alternative methodology has been developed which enables exact results to be determined. This new procedure is described in detail in the paper and its efficacy demonstrated across a range of illustrative and real data sets.

MULTISTAGE PORTFOLIO OPTIMIZATION

Cristinca FULGA
Academy Of Economic Studies, Romania
E-mail: fulga@csie.ase.ro

We consider the problem of a decision maker who is concerned with the management of a dynamic portfolio over a finite horizon. The portfolio optimization problem has a dynamic structure that involves portfolio rebalancing decisions in response to new information on market future prices of the risky assets. Rebalancing decisions are manifested in a sequence of successive revisions of holdings through sales and purchases of assets. We assume that the assets are sufficiently liquid that market impacts can be neglected. We choose the utility function approach to capture the decision maker's risk bearing preferences. The optimal portfolios are obtained using combined techniques based on Progressive Hedging Algorithm (PHA) and Approximate Dynamic Programming (ADP). We obtain the tradeoffs between the expected return and the Conditional Value at Risk (CVaR) measure of losses during the planning horizon from a multi-objective setting, in the Goal Programming (GP) frame. To illustrate the behaviour of the proposed method, we give some significant computational results.

A PHASE TYPE SURVIVAL TREE MODEL FOR CLUSTERING PATIENTS' HOSPITAL LENGTH OF STAY

Lalit GARG, Sally MCCLEAN, Brian MEENAN, Peter MILLARD
University of Ulster, Northern Ireland, UK
E-mail: garg-l@email.ulster.ac.uk

Clinical investigators, health professionals and managers are often interested in developing criteria for clustering patients into clinically meaningful groups according to their expected length of stay. In this paper,

we propose phase-type survival trees which extend previous work on exponential survival trees. The trees are used to cluster the patients with respect to length of stay where partitioning is based on covariates such as gender, age at the time of admission and primary diagnosis code. Likelihood ratio tests are used to determine optimal partitions. The approach is illustrated using nationwide data available from the English Hospital Episode Statistics (HES) database on stroke-related patients, aged 65 years and over, who were discharged from English hospitals over a 1-year period.

IMPLEMENTATION SOLUTIONS AND ISSUES IN BUILDING A PERSONAL SENSOR NETWORK FOR HEALTH CARE MONITORING

Paola Di GIACOMO, Leonardo BOCCHI
University of Udine, Italy
E-mail: paola.digiacomo@ericsson.com

This paper presents a distributed virtual reality environment for children called EVE - Environnements Virtuels pour Enfants. The virtual environment architecture is reactive agents based. The FCM - like dynamic action planning mechanism assures agents' adaptability to its environment changes. This virtual environment supports cooperation among members of a dispersed team engaged in a concurrent context. By the means of their avatars, special cases of agents, users are allowed to interact and to give decisions using cooperative mechanisms. A user-friendly interface enables teachers to create stories that fit with children's pedagogical requirements and generate new virtual environments according to the teacher's specifications. The implementation is based on DeepMatrix as environment server, VRML and Java as languages and Cortona VRML plug-in from ParallelGraphics.

APPLICATION OF STATISTICAL SHAPE ANALYSIS TO THE CLASSIFICATION OF RENAL TUMOURS APPEARING IN EARLY CHILDHOOD

Stefan GIEBEL, Jens-Peter SCHENK, Jang SCHILTZ
University of Luxembourg, Luxembourg
E-mail: jang.schiltz@uni.lu

Most of the renal tumours affecting little children are Wilms-tumours. It is important but very difficult to differentiate Wilms-tumours from other renal tumour types like neuroblastoma, renal cell carcinoma etc. since the correct therapy depends on the diagnosis. For the diagnosis the radiologist has only the magnetic resonance imaging. In many cases the histology by extraction of body tissues is needed additionally.

We try to establish a way to obtain a differential diagnosis using statistical shape analysis on the basis of magnetic resonance which delivers two-dimensional images of different sizes and varying quality. On the basis of magnetic resonance imaging a three-dimensional model of renal tumour is constructed. Then the centre of mass is calculated and the two-dimensional image of the area around the centre of mass is used for shape analysis. Landmarks are computed as the intersection of the boundary of the tumour and a straight line in a pre-determined angle from the centre of mass. The landmarks describe the shape of the tumours. To get a pre-shape the figure has to be centred and standardised. This allows to compare tumours with different sizes situated in different locations. The mean shape is the expected pre-shape of a group of objects/tumours. The mean shape of a group of tumours with a same diagnosis should be able to separate them from tumours with another diagnosis. Tumours with another diagnosis should have a greater distance from the mean shape than the tumours whereby the mean shape is calculated. The applicability of the mean shape for separating groups with different diagnosis is tested by the test of Ziezold (1994). Finally, we present the three-dimensional landmarks and tests for identifying relevant landmarks to separate the different types of tumours.

HIV EVOLUTION THROUGH TWO DIFFERENT TEMPORAL SCALES ACCORDING TO NON-HOMOGENEOUS SEMI-MARKOV MODELS

Di Biase GIUSEPPE, D'Amico GUGLIELMO, Jannssen JACQUES, Manca RAIMONDO
University G. D'Annunzio of Chieti-Pescara, Italy
E-mail: dibiase@unich.it

This paper offers an analysis of HIV/AIDS dynamics, defined by CD4 levels and Viral load, carried out from a macroscopic point of view by means of three different stochastic models. The first model focuses on the patients age as a relevant factor to forecast the transitions among the different levels of seriousness of the disease. The second one is able to make the analysis by considering the disease evolution based on the chronological time. The third model is the most powerful because it considers the two previous features simultaneously. The analyses have been performed through non-homogeneous semi-Markov processes. Such processes have been implemented by using real data provided by ISS (Istituto Superiore di Sanite, Rome, Italy). They refer to 2159 subjects enrolled in Italian public structures from September 1983 to January 2006. The relevant results include also the survival analysis of the infected patients. The computed conditional probabilities show the different responses of the subjects depending on their ages and the elapsing of time. These results allow the measurement of the medical scientific progresses due to the advances in the treatment of the HIV.

ON TYPES OF ALMOST PERIODICITY FOR RANDOM SIGNALS

Ioan GOLET
"Politehnica" University of Timisoara, Romania
E-mail: io_goro@yahoo.co.uk

Almost all naturally occurring phenomena are non-deterministic, that is, they are not entirely predictable. Non deterministic signals are often called random, probabilistic or stochastic. In communication, radar, electrical circuits, computer and telephone networks and many other cases, most signals are random in nature. This is mainly due to the unavoidable omnipresence of different types of noises. Some approximation of random signals are obtained by sampling at different time moments. Time series obtained by such measuring processes show either a combination of periodic phenomena with stochastic components or chaotic behavior. Usually, the computing of nonlinear characteristics indicates the real complexity of the system. In many cases, the separation of frequency bands representing periodic or almost periodic behaviors, allows comprehension of the hidden nonlinear or stochastic phenomena involved.

In this paper we give some results about the approximation of continuous random signals and we study some types of almost periodicity of continuous signals based on probabilistic distributions at the sampling moments.

ADAPTIVE REDUCED-BIAS TAIL INDEX AND VALUE-AT-RISK ESTIMATION

M. Ivette GOMES, Sandra MENDONCA, Dinis PESTANA
Universidade de Lisboa, Portugal
E-mail: ivette.gomes@fc.ul.pt

In this paper, we deal with the estimation, under a semi-parametric framework, of a positive extreme value index, the primary parameter in Statistics of Extremes, and associated estimation of the Value at Risk (VaR) at a level p , the size of the loss occurred with fixed small probability p . Under such a context, the classical extreme value index estimator is the Hill estimator (Hill, 1975), based on an intermediate number k of top order statistics. But the Hill estimator and the associated VaR-estimator lead to sample paths with a high variance for small k , a high bias for large k , and a very sharp mean squared error pattern, as function of k . Recently, several authors have been dealing with bias reduction in the field of extremes and a simple class of second-order reduced-bias extreme value index estimators is the one in Caeiro et al. (2005), used for a semi-parametric estimation of VaR in Gomes and Pestana (2007). We shall here consider those second order

minimum variance reduced bias estimators, and propose the use of bootstrap computer-intensive methods for the adaptive choice of the threshold. An application in the field of insurance will be provided.

References:

- [1] Caeiro, F., Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* 3 (2), 111-136.
- [2] Gomes, M.I. and Pestana, D. (2007). A study reduced-bias extreme quantile (VaR) estimator. *J. American Statistical Association*, Vol. 102, no. 477, 280--292.
- [3] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3, 1163-1174.

EFFECTS OF EXTREME TEMPERATURES ON MORTALITY: A STUDY OF TWO DIFFERENT DISTRICTS OF PORTUGAL

Dulce GOMES, Sandra LAGARTO, Carla NUNES
CIMA - University of Evora, Portugal
E-mail: dmog@uevora.pt

Seasonal fluctuations on mortality are being associated with climate changes and extreme temperatures. Nowadays it's getting more and more important to understand the influence of not only high but also low temperatures on both excess summer/winter deaths in order to evaluate the impacts of these variables on mortality. According to National Directorate of Health (DGS)) these impacts take special relevance within some population groups, like infant and the elder, who are more expose to weather conditions. Research developed in Portugal suggests that, in some regions, for example, with long and hot summer and mild winters, there seems to be some a progressive adaptation that reduces the number of deaths related with heat waves (comparatively with others where the same weather conditions are considered extremely adverse for local populations). In this study, to analyse the effect of extreme temperatures on mortality, we choose 2 different districts - Évora and Bragança (the first in the South and the second in the North-Eastern Portugal) that have different weather patterns. Available data comprises minimum and maximum daily temperatures between 1980 and 1997 and daily mortality. Cross correlation is applied between number of deaths and covariates daily maximum temperatures and minimum temperatures.

ON CONSISTENT CRITERIA FOR NOT-CONSISTENT ALTERNATIVES

Alexander GRUSHO, Elena TIMONINA, Zeev VOLKOVICH
Moscow State University, Russia
E-mail: grusho@yandex.ru

In this paper we consider the existence of consistent criteria for hypothesis testing according to non-consistent probability measures. Such problems arise in many classification tasks like intrusion and abnormal behaviour detections. Here, simple hypotheses $H_{0,n} : P_{0,n}$ are tested against general alternatives $H_{1,n} : \{P_{1,\theta,n} : \theta \in \Theta\}$ in the case where the distributions are defined on the sets X^n , $n = 1, 2, \dots$, produced by a finite set X . These probability measures are also considered on the algebras of the cylindrical sets in the infinite sequences space. We concentrate on the non-consistent families $P_{0,n}$ and $\{P_{1,\theta,n} : \theta \in \Theta\}$. Conditions for different understanding of consistent sequence of tests are investigated.

ABOUT THE LARGEST SUBTREE COMMON TO SEVERAL PHYLOGENETIC TREES

Alain GUENOCHÉ, Henri GARRETA, Laurent TICHIT
CNRS, Marseille, France
E-mail: guenoche@iml.univ-mrs.fr

Given several phylogenetic trees on the same set of taxa X , we look for a largest subset Y in X such that all the partial trees reduced by Y are identical. The problem has polynomial complexity when there are only two trees but it is NP-hard for more than two. We introduce a polynomial approximation algorithm for the multiple problem, which is easy to implement, very efficient and which produces a maximal subtree. It begins by computing an upper bound for the maximum common subtree size and designates elements in X that cannot belong to a common subtree of a given size. Simulations on random and real data have shown that this heuristic always get an optimal solution as soon as the number of taxa is lower than 100. Then, we develop a statistical study to determine the critical size of a MAST to be significant.

RADIUS-MARGIN BOUND ON THE LEAVE-ONE-OUT ERROR OF THE LLW-M-SVM

Yann GUERMEUR, Emmanuel MONFRINI
LORIA-CNRS, France
E-mail: Yann.Guermeur@loria.fr

Using a support vector machine (SVM) requires to set the values of two types of hyperparameters: the soft margin parameter C and the parameters of the kernel. To perform this model selection task, the method of choice is cross-validation. Its leave-one-out variant is known to produce an estimator of the generalization error which is almost unbiased. Its major drawback rests in its time requirement. To overcome this difficulty, several upper bounds on the leave-one-out error of the pattern recognition SVM have been derived. Among those bounds, the most popular one is probably the radius-margin bound. In this article, we establish a generalized radius-margin bound dedicated to the multi-class SVM of Lee, Lin and Wahba (LLW-M-SVM).

A NEW GOODNESS OF FIT TEST FOR THE LOGISTIC DISTRIBUTION

Sneh GULATI, Samuel SHAPIRO
Florida International University, USA
E-mail: gulati@fiu.edu

The logistic distribution is often used in survival analysis and biological studies to model growth curves. In this paper we present a method to test if the available data come from a logistic distribution. The proposed test is a modification of the Greenwood statistic based on higher order spacings as proposed by Rao and Kuo (1984). The percentiles of the test statistic can be approximated by a linear function of the sample size. Power comparisons of the test are carried out via simulations.

ON SOME ESTIMATORS UNDER A MULTIVARIATE MULTIPLE LINEAR MODEL AND ASSOCIATED REDUCED MODEL

Nesrin GULER, Halim OZDEMIR, Mujgan TEZ
Department of Mathematics, Sakarya University, Sakarya, Turkey
E-mail: d045016051@sakarya.edu.tr

In this study, a multivariate multiple partitioned linear model where the model matrix may be deficient in rank and the variance-covariance matrix is nonnegative definite, and a corresponding reduced linear model are considered. Particularly under a condition, we are interested in the case where the best linear unbiased estimator (BLUE) for the expectation of the observable random matrix under the reduced linear model is always an admissible estimator under the partitioned linear model. We also consider an alternative linear estimator and its coincidence with the BLUE under the partitioned linear model. We extend some results in

the literature for the multivariate linear model related to admissible estimator and alternative linear estimator to those in the multivariate multiple case.

MODELING THE SPREAD OF PANDEMIC INFLUENZA FOR INSURANCE INDUSTRY

Marine HABART, Philippe LENCA, Alain HILLION
TELECOM Bretagne, France
E-mail: marine.habart@gmail.com

As far as Insurance Industry is concerned, the emergence of a pandemic could obviously lead to an excess of death in the insured population, and therefore to a possible bankruptcy. Consequently, anticipating and quantifying this risk has become of paramount importance, and this is totally consistent with Solvency II, which is the new set of regulatory requirements for insurance firms that operate in the European Union. This article aims to build a stochastic epidemic model to forecast the spread of pandemic influenza. Initially, a small number of infected people appear in one city, and infection is spread to other cities when they travel. The model includes the effects of prevention and control strategies. Moreover, this model is well adapted to insurance companies because it not only estimates the number of dead people, but also identifies the number of dead insured people, and therefore the potential cost for insurance companies.

STABILITY CONDITIONS OF A PREEMPTIVE REPEAT PRIORITY QUEUE WITH CUSTOMER TRANSFERS

Qi-Ming HE, Jingui XIE, Xiaobo ZHAO
Dalhousie University, Canada
E-mail: qi-ming.he@dal.ca

This paper is concerned with the stability of a preemptive repeat priority queueing system with multiple types of customers and customer transfers between queues. Using the mean-drift method and matrix-analytic methods, simple conditions for the queueing system to be stable or instable are found. The conditions indicate that system stability depends only on the service rate of customers of the highest priority and arrival rates of individual types of customers. That implies that the service rates and the transfer rates of all lower priority customers have no impact on system stability, a result that can be useful in the design of such queueing systems.

HETEROGENEOUS STORE-LEVEL SALES RESPONSE MODELS WITH ENDOGENOUS STORE ATTRIBUTES

Harald HRUSCHKA
University of Regensburg, Germany
E-mail: harald.hruschka@wiwi.uni-regensburg.de

The number of papers studying store-level sales response models which consider endogeneity appears to be very limited. We consider potential endogeneity of store attributes in the sales response function by an instrumental variable approach. We also allow for heterogeneity across stores by assuming that store-level coefficients are generated by a finite mixture distribution. Models are estimated by a MCMC simulation technique which combines two Gibbs sampling algorithms. In the empirical study both heterogeneity and endogeneity turn out to influence estimates. For a cross section of more than 1,000 gas stations credible intervals of differences of coefficients between models ignoring and models considering endogeneity are computed. These intervals indicate that models which ignore endogeneity overestimate the effect of two store attributes on sales.

A PARAMETRIC METHOD FOR PATTERN CLASSIFICATION

Iuliana IATAN
Technical University of Civil Engineering Bucharest, Romania
E-mail: iuliafi@yahoo.com

In this paper I have generalized the test likelihood ratio for multivariate normal distribution. The aim of this paper is the analysis of this supervised parametric classification method, based on some modern methods of multidimensional statistical analysis. The power of the test is calculated for our proposed parametric classification criterion.

THE EXPECTATION – MAXIMIZATION ALGORITHM: GAUSSIAN CASE

Iuliana IATAN
Department of Mathematics and Informatics,
Technical University of Civil Engineering Bucharest
E-mail: iuliafi@yahoo.com

There are some situations when in the pattern recognition applications can appear some objects which are missing data. This thing one happens since the process of data acquisition isn't perfect. In this paper we shall present the EM algorithm (Expectation- Maximization) which is used in order to estimate the parameters corresponding to a probability density function when we dispose by missing data. In our case, the class labels are the missing data.

TESTING GOODNESS-OF-FIT OF HIGH-DIMENSIONAL GAUSSIAN MIXTURES

Gintautas JAKIMAUSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: gnt@ktl.mii.lt

Let us have a sample satisfying d -dimensional Gaussian mixture model (d is supposed to be large). The problem of classification of the sample is considered. Because of large dimension it is natural to project the sample to k -dimensional ($k = 1, 2, \dots$) linear subspaces using projection pursuit method (Huber (1985), Friedman (1987)) which gives the best selection of these subspaces. Having an estimate of the discriminant subspace we can perform classification using projected sample thus avoiding 'curse of dimensionality'. An essential step in this method is testing goodness-of-fit of the estimated d -dimensional model assuming that distribution on the complement space is standard Gaussian.

We present a simple, data-driven and computationally efficient procedure for testing goodness-of-fit. The procedure is based on well-known interpretation of testing goodness-of-fit as the classification problem, a special sequential data partition procedure, randomization and resampling, elements of sequential testing. Monte-Carlo simulations are used to assess the performance of the procedure.

OPTIMAL SAMPLING SIZE ALLOCATION FOR MULTI-LEVEL STRESS TESTING WITH EXTREME VALUE REGRESSION

C.Y. KA, P.S. CHAN, H.K. NG, N. BALAKRISHNAN
The Chinese University of Hong Kong
E-mail: benchan@cuhk.edu.hk

In this talk, we discuss the optimal allocation problem in a multi-level stress testing experiment under Type-II censoring when an extreme-value regression model is used for statistical analysis. We derive the maximum likelihood estimators, the Fisher information, and the asymptotic variance-covariance matrix of the maximum likelihood estimators. Three optimality criteria are used to discuss the optimal allocation problem. Optimal allocation of units, both exactly for small sample sizes and asymptotically for large sample

sizes, for two- and four- level situations are determined numerically. Conclusions and discussions are provided based on the numerical studies.

ON SIMULATION OF RUIN USING INTEGRATED TAIL DISTRIBUTION

Ants KAASIK, Kalev PARNA
University of Tartu, Estonia
E-mail: ants.kaasik@ut.ee

In this paper we study the approximation of the tail of the integrated tail distribution (itd) when we have data available only from the original distribution. The approximation is relevant when we wish to estimate the infinite time ruin probability of a Cramer-Lundberg insurance risk process using only claim data and no additional information. Main tool for the task is the generalized Pareto (GP) distribution. We discuss why our approach leads naturally to the GP estimator of the expected value (introduced in Johansson, 2003). Automatic threshold selection is used in the simulation study, making better use of a specific sample. Comparison of the approximations is split into two intervals - one within the sample range and the other outside of it - allowing more interesting conclusions to be drawn. This way we extend the results obtained in Kaasik (2008).

MAD APPROACH FOR PORTFOLIO MANAGEMENT WITH STABLE ASSETS

Audrius KABASINSKAS, Igoris BELOVAS, Leonidas SAKALAIUSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: audrius.kabasinskas@ktu.lt

Construction of an investment portfolio is discussed in many papers, but, traditionally when there is no assumption on data series distribution, the standard deviation or variance is assumed as a risk measure. During the last few years historical data series distribution has often been analyzed and new ways of portfolio management proposed. It has been shown that financial data are stable or mixed-stable (in emerging or FOREX markets) distributed. In such cases new relation measures and portfolio construction techniques are required. In this paper we focus on the portfolio construction problem under the assumption of data stability or mixed-stability.

BUSY PERIOD, TIME OF THE FIRST LOSS OF A CUSTOMER AND THE NUMBER OF CUSTOMERS

Tetyana KADANKOVA, Noel VERAVERBEKE
Hasselt University, Belgium
E-mail: tetyana.kadankova@uhasselt.be

Queueing systems with finite waiting room have wide applications in the performance evaluation, telecommunications, and manufacturing systems. The aim of our work to derive the most important performance measurements of several queueing systems of this type. More precisely, we consider the $M^k/G^\delta/1/B$ and $G^\delta/M^k/1/B$ queueing systems with finite waiting room and their modifications. Evolution of the number of the customers in such systems is described by a process with two reflecting boundaries. In general case this process is a difference of two renewal processes. Reflections from the upper boundary are generated by the supremum (infimum) of the process. Reflections from the lower boundary govern the server's behavior. In general such processes are not Markovians, but by adding a complementary linear component (in some literature called age process), we obtain a Markov process, which describes functioning of the queueing system. The workload process of the finite capacity queue results to the difference of the two compound renewal process. Studying main characteristics of the system results to the investigating of the two-boundary functionals of the governing process. For the queueing systems of $M^k/G^\delta/1/B$, $G^\delta/M^k/1/B$ type the governing process is the difference of the compound Poisson process and the compound renewal process complemented with the age process. We determine the Laplace

transforms of busy period, time of the first loss of the customer and the number of the customers in the system. Additionally, we consider a special case, when the governing process has unit negative jumps. It means that the customers arrive not in batches but one-by-one. In this case we derive more tractable results. We also study the $G^{\delta}/M^k/1/B$ system, for which we determine the busy period, time of the first loss, distribution of the number of the customers in the system and the virtual waiting time. Partial case is treated separately.

APPLICATION OF CHAOS THEORY TO ANALYSIS OF COMPUTER NETWORK TRAFFIC

Liudvikas KAKLAUSKAS, Leonidas SAKALAUSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: liudas@fm.su.lt

Article deals with statistical university network traffic, by applying the methods of self-similarity and chaos analysis. The object of measurement is Siauliai University LitNet network node maintaining institutions of education of the northern Lithuania region. Time queues of network traffic characteristics are formed by registering amount of information packets in a node at different regimes of network traffic and different values of discretion of registered information are present. Measurement results are processed by calculating the correlation measure of the queue, analysing attractors, estimating the fractal measure, calculating Herst index and estimating reliability of analysis results by applying the statistical method. Investigation of the network traffic allowed us drawing conclusions that time queues bear features of self-similarity when aggregated time queues bear features of slowly decreasing dependence, and time series' attractors depend on the group of strange attractors.

ALTERNATING RENEWAL PROCESSES WITH CORRELATED COMPONENTS

Revaz KAKUBAVA
Georgian Technical University, Georgia
E-mail: r.kakubava@gmail.com

Residual life-time and residual renewal (repair) time are the important characteristics of renewable systems [1,3,4]. Analysis of such characteristics is carried out within the framework of the renewal theory [1]. However, when life-time and renewal time are interdependent (correlated), probabilistic analysis of residual life-time and residual renewal time by the methods of classical renewal theory is difficult, if not impossible. At the same time the technical systems, in which there is no possibility for negligence of such correlation are widespread.

In the presented work two kinds of such systems are described and the alternating renewal process in which the life-time and the renewal time are interdependent is examined. Probabilistic characteristics of such processes in the terms of Laplace transforms are obtained.

Undoubtedly these results have a wide practical applications.

References:

1. Beichelt, F., Franken, P. (1988): Nadezhnost i technicheskoe obsluzhivanie. (In Russian). Ed. I. Ushakov. Moscow: Sovetskoe Radio.
2. Ushakov, I. (2006): Reliability: Past, Present, Future. Electronic Journal. Reliability: Theory and Applications. Vol.1, No.1. <http://gnedenko-forum.org/Journal/>.
3. Kakubava, R.V., Khurodze, R.A.(2000): Probabilistic analysis of a duplicated system with recovery and switching. Automation and Remote Control, Vol. 61, No. 9, Part 1
4. Kakubava.R.,(2008) Analysis of Alternating Renewal Processes with Depended Components. Reliability and Risk Analysis: Theory and Applications. No.1, Vol.1. <http://gnedenko-forum.org/Journal>.

- 5 Goel L.R., Shrivastana P. (1991): A two-unit cold standby system with three modes and correlated failures and repairs. *Microelectron. Reliab*, Vol. 31, No.5.
6. Goel, L.R., Mumtaz, S.Z, Gupta, R.(1996): A two-unit duplicating standby system with correlated failure-repair/replacement times. *Microelectron. Reliab*, Vol. 36, No.4.

THE NEXT GENERATION OF DATA ANALYSIS TOOLS

Alexandros KARAKOS, Pericles KARAKOS
Democritus University of Thrace, Greece
E-mail: karakos@ee.duth.gr

Purpose of this paper is to investigate on the next generation of data analysis tools.

Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable. When available information is stored in database tables containing rows and columns, Multivariate Analysis can be used to process the information in a meaningful fashion.

AJAX is a new implementation of established web development technologies to gain interactivity between users and servers through multiple client sided server connections in the background. AJAX is a kind of framework that leads Web programming into the direction of Web 2.0, the new generation of using and developing the internet. With Ajax, Web applications finally start feeling like desktop applications to the Internet users. The use of Ajax programming in the field of data analysis can to investigate on the next generation of data analysis tools. As example, to demonstrate our approach to the new generation of data analysis tools, we present PRAXITELE, an online service for applications of data analysis. You can find PRAXITELE at: <http://peramos.ee.duth.gr/dataAnalysis2/dataAnalysis/praxitelis/English/index.html>. This service provides to the users a friendly environment, where they can manage their data and apply on line different methods of Multivariate Data Analysis.

DEPENDENCE OF THE LAPLACIAN EIGENMAPS METHOD AND ITS MODIFICATION ON THE PARAMETERS

Rasa KARBAUSKAITE, Gintautas DZEMYDA
Institute of Mathematics and Informatics, Lithuania
E-mail: karbauskaite@ktl.mii.lt

This paper deals with a method, called Laplacian Eigenmaps (LE). It is a nonlinear dimensionality reduction technique and it finds a low-dimensional data representation by preserving local properties of the manifold. The modification of the Laplacian Eigenmaps algorithm proposed in this paper differs from the original LE algorithm by parameters that influence the mapping quality. In the modified LE algorithm, selection of the parameters is simpler. The variable number of the nearest neighbours for each data point is a uniqueness of the modification.

ON UNIFORMLY OPTIMAL CONTROL IN AN INFORMATION THEORY

Pavel KATYSHEV
CEMI RAN, Moscow, Russia
E-mail: pkatish@nes.ru

The problem of a transmission of the Gaussian messages through the white Gaussian channel with feedback is considered as a problem of the control of the Ito process. The objective function to be minimized is the mean square error. The class of the transmitted processes is found for which the uniformly optimal control (coding and decoding scheme) exists. For the Gaussian and Markovian transmitted process the optimal value of the objective function is calculated in terms of the covariance function of the underline process.

PROBABILISTIC MODELLING OF SOIL WATER IN A SEMIARID BANDED VEGETATION SYSTEM

Laimonis KAVALIERIS, Tim ELLIS
University of Otago, Dunedin, New Zealand, Tim
E-mail: lkavaliaris@maths.otago.ac.nz

Banded vegetation systems occur naturally in many semiarid environments. They typically comprise alternate runoff and runoff-on zones, oriented across a hillslope and replicated downslope. Bare, crusted zones with low infiltration capacity, generate overland flow (runoff), which then cascades onto vegetated zones of high infiltration capacity. In this way water supply to the vegetation is increased above what is delivered via rainfall alone. The banded mulga (*Acacia aneura*) systems in Australia comprise three zones: bare soil; grass; and trees, with infiltration and water storage capacities increasing roughly stepwise, and in that order via runoff onto the next zone. It has been proposed that agroforestry systems can be designed to make use of this passive method of water redistribution, allowing bands of productive trees to be grown in environments with relatively low rainfall.

Daily rainfall was modelled as a marked Poisson process; the soil water storage as a finite storage defined by soil porosity and plant root depth; and evapotranspiration as a loss function dependent on the level of soil water storage. Each of the three zones can be modelled independently, but here they are modelled as a coupled stochastic system. We derive stationary distributions for soil water storage and surface runoff. Future work will include application of this model to natural and agricultural systems.

OUTLIER COMPONENTS FOR TIME SERIES MODELS

Ahmet KAYA, Nurkut Nuray URGAN
Namik Kemal University, Turkey
E-mail: nurgan@nku.edu.tr

Error in data is one of the facts that cause the parameter estimations to be subjective. If the erroneous case is proved statistically, then these cases are called outliers. Outliers may occur either as a result of the natural randomness or because of the mistakes of data operators or equipments. In this study, a multi-factor ($2 \times 3 \times 2$) modelling is done in order to fix the effects of model in data analysis (AR, MA), generosity of the data (3-sigma, 5-sigma), serie range (50, 100, 150) and criterion value sensibility (% 99 ($C=3.00$), % 95 ($C=3.50$), % 90 ($C=4.00$)) factors statistically by making use of a simulation results made by Chang Tiao and Chen.

ON ONE APPROACH TO STUDY OF DEPARTURE PROCESS IN BATCH ARRIVAL QUEUES WITH MULTIPLE VACATIONS

Wojciech KEMPA
Silesian University of Technology, Poland
E-mail: Wojciech.Kempa@polsl.pl

A batch arrival single-server queueing system of a general type is considered. At the end of each busy period the server begins a multiple vacation period when the service process is stopped. A new approach for the study of departure process in such a system is proposed. Using the formula of total probability and renewal theory we direct the analysis to that in the system without vacations. An explicit representation for the double transform of probability function of departure process is derived. The formula if written down by means of transforms of distributions describing the arrival and service processes and components of a certain factorization identity of Wiener-Hopf type connected with these distributions.

BOUNDARY PROBLEMS FOR STOCHASTIC EQUATIONS AND STOCHASTIC MARKOVIAN FIELDS

Khaled KHALDI, Yamina BOUZAGHTI
Boumerdes University, Algeria
E-mail: kkhaldi@umbb.dz

Some models of probabilities are decrypted by generalized stochastic equations. These models (such that of the prediction) lead to the resolution of boundary problems for random generalized equations. We are interested in the equation $L\xi = \eta$ in an open $S \subseteq R^d$, where L is a linear operator, and ξ and η are random distributions and with the class of boundary conditions $\Gamma \supseteq \partial S$ which it is necessary to define for the corresponding boundary conditions.

ROBUSTNESS IN STATISTICAL FORECASTING OF TIME SERIES

Yuriy KHARIN
Belarusian State University, Minsk
E-mail: kharin@bsu.by

Optimality of the majority of the known statistical forecasting procedures is proved w.r.t. the mean square risk of prediction under the assumptions of the underlying hypothetical model. In practice, however, the observed time series satisfy the hypothetical models with some distortions. Unfortunately, the forecasting procedures, which are optimal under hypothetical models, often lose their optimality and become unstable under "a little distorted hypothetical models". The following main problems considered in this lecture are very topical: A) robustness evaluation of the traditionally used statistical forecasting procedures under distortions; B) evaluation of delta-admissible distortion levels; C) construction of new robust forecasting procedures by the minimax risk criterion. In the lecture we give solutions of the problems A, B, C for the following cases: trend models of time series under "outliers" and functional distortions of trends; regression models under "interval" distortions, "relative" distortions and distortions in the L_p -metrics; autoregressive time series under non-homogeneities in the innovation process and under bilinearities; vector autoregressive time series under "misspecifications"; vector autoregressive time series under missing values.

Theoretical results are illustrated by the results of computer experiments on simulated and real time series.

RISK MODELS WITH EXTREMAL SUBEXPONENTIALITY

Dimitrios KONSTANTINIDES
University of the Aegean
E-mail: konstant@aegean.gr

In this paper we consider risk models with a heavy-tailed parametric claim distribution from the subexponential class \mathcal{S} with at least two parameters. We choose a proper convergence of a parameter, that makes the tail of the claims distribution heavier or lighter and then tend it to its limitation. Finally we proceed to an appropriate functional normalization in order to keep the distributional properties.

MODELLING OF THE MULTIPRODUCT INVENTORY PROBLEM WITH RESOURCES LIMITATION

Eugene KOPYTOV, Leonid GREENGLAZ, Aivars MURAVJOVS
Transport and Telecommunication Institute
E-mail: kopitov@tsi.lv

The given paper considers a multiproduct inventory control model with constraints, which determine restrictions on the resources (stock capacity, capital and others). We propose the model with a fixed period of

time between the moments of placing neighbouring orders, and the order quantity is determined as the difference between the fixed stock level and the quantity of goods at the moment of ordering. It is assumed that each order may enclose only fixed kinds of products. In the considered problem demands for products are random with known distributions and the lead time is constant.

The process of the task solving is divided into two stages:

- search of the starting point (an approximate solution of the inventory control deterministic problem and the correction of the received solution with the account of losses from the possible goods deficit and the random demand distributions);

- finding an optimal solution of the inventory problem using the simulation approach.

The considered model is realized by the combination of the analytical and simulation approaches in packages MS Excel and ExtendSim correspondingly. The numerical examples of problem solving are presented in the paper. The considered model is of interest because it illustrates a real business situation.

THE TRANSITION FROM UNIVERSITY TO FIRST EMPLOYMENT: A NON PARAMETRIC SURVIVAL APPROACH OF THE GREEK SOCIAL SCIENCE GRADUATES

Foteini KOUGIOUMOUTZAKI, Dimitios KALAMARAS, Aglaia KALAMATIANOU
Panteion University of Athens, Department of Sociology
E-mail: dimitris.kalamaras@panteion.gr

The expansion of higher education and the changes in the world of work during the last few decades, have strongly affected the relationship between higher education and employment. One aspect of this relationship that has been considerably affected concerns the transition process from higher education to the labour market, which nowadays is long and complex. The relevant studies aim to reveal challenges and problems graduates are facing in their effort to begin career life. Attention is drawn to the analysis of the transition time between graduation and first employment as well as to the factors related. This paper focuses on this issue for the case of Greece; a country with high proportion of tertiary level students and high graduate unemployment rates, comparatively to other European Union countries. In particular, we examine the time duration to obtain the first job and the factors related, using a sample of 688 university graduates oriented to social and political sciences. The analysis employs a non parametric survival technique, the Kaplan-Meier estimator. The results revealed a rather long job search period emerging difficulties in getting a job, and a number of factors affecting this period.

ON ESTIMATION AND ASYMTOTICS OF THE HURST INDEX OF SOLUTIONS OF STOCHASTIC INTEGRAL EQUATIONS

Kestutis KUBILIUS, Dmitrij MELICHOV
Institute of Mathematics and Informatics, Lithuania
E-mail: kubilius@ktl.mii.lt

We consider the solutions of stochastic integral equations driven by the fractional Brownian motion with the Hurst index H . Studying the asymptotic behavior for the first and second order quadratic variation of the solutions of stochastic equations, we obtain strongly consistent estimators of the Hurst index H . We shall prove that the estimators are asymptotically normal.

SURVIVAL PROBABILITY ANALYSIS OF DETERIORATING MEMBERS AND SYSTEMS

Antanas KUDZYS, Ona LUKOSEVICIENE
Kaunas University of Technology, Kaunas, Lithuania
E-mail: olukoseviciene@gmail.com

Introduction. The attention of reseachers and engineers to the merits of probability-based methods in the prediction of durability of deteriorating members and systems is discussed.

Structural safety margin of members. The stochastic time-dependent safety margin of deteriorating members exposed to permanent and transient extreme action effects is modeled. These effects are treated as random rectangular pulse renewal processes. In design practice, the safety margin process is modeled as a random geometric distribution and treated as a finite random sequence.

Survival probability prediction. The correlation factor of survival probabilities of deteriorating systems cuts is analysed. Instantaneous and long-term survival probabilities of deteriorating members are considered. The model of time-dependent safety prediction of these members is based on approaches of transformed conditional probabilities.

Reliability index of deteriorating members. The initial and revised long-term reliability indices of deteriorating members are presented. Their values are calculated in a simplified but quite exact analysis manner. It is illustrated by numerical example.

Conclusion. It is expedient to base in design practice the prediction of time-dependent reliability and survival probability of deteriorating members on the method of transformed conditional probabilities.

PROBABILISTIC MODELLING AND SENSITIVITY ANALYSIS OF THE AIRCRAFT CRASH CONSIDERING UNCERTAIN DATA

Kristina Kupciuniene, Robertas Alzbutas
Lithuanian Power Institute, Lithuania
E-mail: kriskup@mail.lei.lt

In this paper the probabilistic analysis of aircraft crash on the strategically important object is performed. Plane or other flying object fall on the territory of the strategically important object represents an extensive danger to it. Thus, the modelling of aircraft crash occurrence is significant in order to evaluate risk of such event and prevent potential failure of the systems important to safety of the object. In this paper, the improved aircraft crash probability estimation model is presented. The aircraft crash probability on the territory of the strategically important object with the specified radius may be expressed by the derived analytical formula. This formula is used to calculate the aircraft crash probability applying the most recent but still uncertain statistical data. The aircraft crash dependencies upon the amount of flights and crashes in the different radius territories were analysed. As a part of the initial conditions and parameters of aircraft crash model are not well-known or have different values for various types of planes, the sensitivity and uncertainty analysis was performed. This analysis determines the parameters that have the greatest influence on the model results. Conclusions about the importance of the parameters and sensitivity of the results are obtained using sensitivity measures of the model under analysis.

VISUALIZATION OF SUPPORT VECTORS

Olga KURASOVA
Institute of Mathematics and Informatics, Lithuania
E-mail: Kurasova@ktl.mii.lt

In this paper, some ways of visualization of support vectors are investigated. We examine a possibility to visualize support vectors by various methods, for example, Chernoff faces, Andrew's curve, parallel coordinates, stars, etc. Since the changes of the classification quality are slight when comparing the results of classification of multidimensional data with that of two-dimensional data obtained by multi-dimensional scaling, it is possible to determine decision surfaces by analyzing the mapping data with a view to get more understandable results.

INVESTIGATION OF QUALITY OF MAPPING VECTORS OBTAINED BY QUANTIZATION METHODS

Olga KURASOVA, Alma MOLYTE
Institute of Mathematics and Informatics, Lithuania
E-mail: alma.molyte@gmail.com

In this paper, a quality of mapping vectors obtained by vector quantization methods (self-organizing map and neural gas) is investigated. Multidimensional scaling is used for mapping. Two criteria (topology preserving measure and Spearman rho) are applied for the estimation of mapping quality.

NON-GAUSSIAN AUTOREGRESSIVE TIME SERIES MODELS

KUTTYKRISHNAN A.P.
Department of Statistics, Sir Syed College, Taliparamba, Kerala, India
E-mail: kutykrishnanap@rediffmail.com

Geometric stable distributions can be used in modeling peaked and heavy tailed observations that may be a result of a random number of independent innovations. Although the theory and applications of geometric stable distributions is well developed and appeared in many literature in recent years, their applications in time series modeling are not developed. Present work aims at developing autoregressive model using geometric stable distributions as marginal distributions.

ADAPTIVE TRANSMISSION FOR LOSSLESS IMAGE RECONSTRUCTION

Elisabeth LAHALLE, Gilles FLEURY, Rawad ZGHEIB
Supelec, France
E-mail: elisabeth.lahalle@supelec.fr

This paper deals with the problem of adaptive digital transmission systems for lossless reconstruction. A new system, based on the principle of non-uniform transmission, is proposed. It uses a recently proposed algorithm for adaptive stable identification and robust reconstruction of AR processes subject to missing data. This algorithm offers at the same time an unbiased estimation of the model's parameters and an optimal reconstruction in the least mean square sense. It is an extension of the RLSL algorithm to the case of missing observations combined with a Kalman filter for the prediction. This algorithm has been extended to 2D signals. The proposed method has been applied for lossless image compression. It has shown an improvement in bit rate transmission compared to the JPEG2000 standard.

REGRESSION MODELS WITH RIGHT-TRUNCATED DATA. APPLICATIONS TO THE ANALYSIS OF THE INCUBATION TIME IN A SAMPLE OF AIDS-DIAGNOSED PATIENTS

Ana LARA-PORRAS, Yolanda ROMAN-MONTOYA, M.Luz GAMIZ
University of Granada, Spain
E-mail: mgamiz@ugr.es

In this work, we analyse a sample of AIDS-diagnosed patients in Andalucia (Spain), with registers from January, 1981 to June, 2008. The interest focuses in the study of the elapsed time (in months) between two main events: the HIV infection (initial event) and the diagnosis of AIDS (final event). We refer this elapsed time as the incubation time. Observations on the incubation time are right truncated since only individuals that experience the final event prior to June, 2008 are taken into account.

The main objective of the work is to evaluate the incidence that several covariates have on the incubation time. The data have been provided by the Registro Andaluz de Casos de Sida and contain

sampling information concerning the following covariates: gender, age at diagnosis, transmission category, disease at AIDS diagnosis, county of residence (hospital), and the type of HAART (if it is the case).

Taking as response the risk function of the incubation time (in months), a semiparametric model is built based on the proportional hazards assumption and adapted to manage right-truncated datasets.

We use R to provide an estimation procedure of the semiparametric regression model under right truncation. Afterward, we evaluate the effect of each covariate over the risk of developing AIDS, which allows us to identify the main risk factors in this particular population of HIV-seroconverters. Finally, we check the hypotheses under which the model is constructed by examining several residual plots.

ROBUST PERFORMANCE HYPOTHESIS TESTING WITH THE SHARPE RATIO

Olivier LEDOIT, Michael WOLF
University of Zurich, Switzerland
E-mail: mwolf@iew.uzh.ch

Applied researchers often test for the difference of the Sharpe ratios of two investment strategies. A very popular tool to this end is the test of Jobson and Korkie (1991), which has been corrected by Memmel (2003). Unfortunately, this test is not valid when returns have tails heavier than the normal distribution or are of time series nature. Instead, we propose the use of robust inference methods. In particular, we suggest to construct a studentized time series bootstrap confidence interval for the difference of the Sharpe ratios and to declare the two ratios different if zero is not contained in the obtained interval. This approach has the advantage that one can simply resample from the observed data as opposed to some null-restricted data. A simulation study demonstrates the improved finite sample performance compared to existing methods. In addition, two applications to real data are provided.

MODELS OF STATISTIC DISTRIBUTIONS OF NONPARAMETRIC GOODNESS-OF-FIT TESTS IN COMPOSITE HYPOTHESES TESTING IN CASE OF DOUBLE EXPONENTIAL LAW

Boris LEMESHKO, Stanislav LEMESHKO
Novosibirsk State Technical University, Russia
E-mail: Lemeshko@fpm.ami.nstu.ru

In this paper are presented more precise results (tables of percentage points and statistic distribution models) for the nonparametric goodness-of-fit tests in testing composite hypotheses using the maximum likelihood estimate (MLE) for double exponential distribution law. Statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulation.

COMMUNITY DETECTION USING INTELLIGENT CLUSTERING TECHNIQUE AND SUB-MATRIX DENSITY ORDERING

Tianzhu LIANG, Kwok Yip SZETO
HKUST, Kwok YIPp, Honk KONG Chinese University of Hong Kong
E-mail: phszeto@ust.hk

Detecting communities in real world networks is an important problem for data analysis in science and engineering. By clustering nodes intelligently, a recursive algorithm is designed to detect community. Since the relabeling of nodes does not alter the topology of the network, the problem of community detection corresponds to the finding of a good labelling of nodes so that the adjacency matrix form blocks. By putting a fictitious interaction between nodes, the relabeling problem becomes one of energy minimization, where the total energy of the network is defined by putting interaction between the labels of nodes so that the clustering of nodes in the same community will decrease the total energy. A greedy algorithm is used for the computation of minimum energy. The method shows efficient detection of community in artificial as well as real world network. The result is illustrated in a tree showing hierarchical structure of communities on the basis of sub-matrix density.

A CRITERION FOR SELECTING THE NUMBER OF SLICES AND THE DIMENSION OF THE MODEL IN SIR AND SAVE APPROACHES

Benoit LIQUET, Jerome SARACCO
University of Bordeaux, France
E-mail: benoit.liquet@isped.u-bordeaux2.fr

Sliced inverse regression (SIR) and related methods were introduced in order to reduce the dimensionality of regression problems. In general semiparametric regression framework, these methods determine linear combinations of a set of explanatory variables X related to the response variable Y , without losing information on the conditional distribution of Y given X . They are based on a "slicing step" in the population and sample versions. They are sensitive to the choice of the number H of slices, and this is particularly true for SIR-II and SAVE methods. At the moment there are no theoretical results nor practical techniques which allows the user to choose an appropriate number of slices. In this talk, we propose an approach based on the quality of the estimation of the effective dimension reduction (EDR) space: the square trace correlation between the true EDR space and its estimate can be used as goodness of estimation. We introduce a naive bootstrap estimation of the square trace correlation criterion to allow selection of an "optimal" number of slices. Moreover, this criterion can also simultaneously select the corresponding suitable dimension K (number of the linear combination of X). From a practical point of view, the choice of these two parameters H and K is essential. We propose a 3D-graphical tool, implemented in R, which can be useful to select the suitable couple (H, K) . In this talk, we focus on the SIR, SIR-II and SAVE methods. We indicate how the proposed criterion can be used in practice. A simulation study is performed to illustrate the behaviour of this approach and the need for selecting properly the number H of slices and the dimension K .

A PROBABILISTIC APPROACH FOR EVALUATING THE SENSITIVITY TO FAKE DATA IN STRUCTURAL EQUATION MODELING

Luigi LOMBARDI, Massimiliano PASTORE
University of Trento, Italy
E-mail: massimiliano.pastore@unipd.it

In this paper we address the issue of evaluating the sensitivity of goodness-of-fit indices in structural equation modeling when fake data are considered in three different factorial models with varying sample sizes ($n= 50, 100$ and 200). The sensitivity evaluation is carried out by means of a simulation procedure which combines a standard Monte Carlo approach and a new probabilistic version of a recent data perturbation procedure called Sample Generation by Replacements (SGR, Lombardi, Pastore and Nucci, 2004). Probabilistic SGR (PSGR) will be used to generate data perturbations based on three different models of faking: fake-uniform, fake-good (deception) and fake-bad (malingering). For each scenario of faking the performance of four very popular goodness-of-fit indices (two absolute indices: GFI, and AGFI; and two incremental indices: CFI and NNFI) will be evaluated.

QUEEN'S CROQUET GAME IN INSURANCE RISK REGULATION

Vsevolod MALINOVSKII
Finance academy & Steklov Mathematical Institute, Russia
E-mail: malinov@orc.ru

This contribution is devoted to the risk theory insight into the problem of asset-liability and solvency adaptive management in insurance under deficient information. The latter means that the intensities of the successive annual claim arrival processes are the random variables which comply with a certain scenario. The outcomes of these random variables remain unknown at the control times which makes the control similar to the Queen's croquet game described by Lewis Carroll. Several adaptive control strategies compensating the inevitable errors due to both pure randomness and incomplete knowledge of the future risk are introduced and their performance is analyzed.

Two standard techniques aimed to evaluate the impact of deficiency of information on the insurer's business are scenario analysis and stress testing. The former considers typical, favorable and unfavorable scenarios of nature. The latter refers to shifting the values of the individual parameters in the model that affects critically the insurer's financial position. Usually, both these techniques apply simulation. Bearing in mind all that, we will accentuate the risk theory-based, analytical approach.

STATISTICAL ANALYSIS OF THE MULTIVARIATE REGRESSION MODELS WITH HETEROGENEOUS STRUCTURE

Vladimir MALUGIN
Belarus State University, Belarus
E-mail: Malugin@bsu.by

The paper is devoted to statistical analysis of the multivariate regression models with heterogeneous structure. The parameters estimation algorithm for the model with autocorrelated innovations is suggested and examined. The consistent «plug-in» decision rule for classification of the multivariate heterogeneous dependent regression observations is presented. The EM-algorithm for the splitting of a mixture of the multivariate regression observations is also suggested. The convergence of this algorithm is proved and the classification error probability is examined on simulated data.

COLLISIONS IN STOCHASTIC MULTI-COMPONENT SYSTEMS

Anatoly MANITA
Moscow State University, Russia
E-mail: manita@mech.math.msu.su

For n-dimensional stochastic system we define a notion of collision between different components. The main goal is the probabilistic analysis of the system in the limit as n goes to infinity. We consider several classes of collision models and obtain some asymptotic results on probability distribution of the number of collisions in these systems.

IMPROVING POST-MINING OF ASSOCIATION RULES WITH ONTOLOGIES

Claudia MARINICA, Fabrice GUILLET
LINA - Ecole polytechnique de l'Universite de Nantes, France
E-mail: claudia.marinica@univ-nantes.fr

In data mining, the discovery of association rules is strongly limited by the huge amount of delivered rules. In this paper, we propose to improve post-processing of association rules by a better integration of user (decision maker) goals and knowledge. On one hand, we integrate a domain ontology associated to data which extends Generalized Association Rules. On the other hand, we generalize General Impressions with Rule Schemas. Ontologies will offer a powerful representation of user knowledge, and rule schemes a more expressive representation of user expectations in term of rules.

MODELLING THE GEOGRAPHIC DEPENDENCE OF RISK IN HOUSEHOLD INSURANCE

Laszlo MARKUS
Eotvos Lorand University, Hungary
E-mail: markus@cs.elte.hu

We analyse here the spatial dependence structure for the counts of certain type of claims occurring in household insurance in Hungary. We build up a Bayesian hierarchical model for the Gaussian Markov Random Field driven non-homogeneous spatial Poisson process of claims counts. Model estimation is

carried out by the MCMC algorithm, by applying sparse matrices technique. The challenge in the implementation of the MCMC algorithm is the very high dimension. We have to reach convergence and mixing in a 3173 dimensional state space. In the very high dimension the density of the length of a normal vector has a very narrow bandwidth, resulting in a very narrow environment around the Gaussian ellipsoid, where the values of the vector are likely. It is a simple geometrical consideration, that if in high dimension we add to a vector another vector of fixed length with uniform random angle, then with high probability it will increase the length of the original vector and decrease it only with a very small probability (unlike in two dimension, where each of these probabilities hardly differ from 1/2). This means, random walk Metropolis updates will lengthen the vector, whereas the precision parameter in the likelihood tries to keep its length in a narrow band. If we were to avoid frequent rejection, most of the updates has to be of very small length in order to remain within the mentioned narrow environment of likely values. This paralyses the mixing of the chain. To allow for longer updates in order to achieve better mixing and faster convergence we suggest a novel approach for updates by radial and angular components.

RECENT DEVELOPMENTS IN FITTING COXIAN PHASE-TYPE DISTRIBUTIONS IN HEALTHCARE

Adele MARSHALL, Mariangela ZENGA
CenSSOR, Northern Ireland
E-mail: a.h.marshall@qub.ac.uk

In the past few decades, Coxian phase-type distributions have become increasingly more popular as a means of representing survival times in healthcare models. In particular, they are considered suitable for modelling the length of stay of patients in hospital and more recently for modelling the patient waiting times in Accident and Emergency Departments. However, problems have arisen in how to accurately estimate the parameters of the model from healthcare data. This paper examines the various approaches and considers the recent developments in fitting the Coxian phase-type distribution.

WRAPPER AGGREGATION FOR CLASSIFICATION

Tristan MARY-HUARD, Stephane ROBIN
AgroParisTech/INRA, France
E-mail: maryhuar@agroparistech.fr

Supervised classification is an important field of investigation that has received a lot of attention from computational scientists and statisticians in the last decade. With the recent development of high-throughput technologies, many strategies have been developed to deal with large dimension data, where the number of observations n is lower than the number of variables p . In this particular context, building an interpretable classifier is difficult, due to the presence of many irrelevant variables, and to a high level of redundancy between the variables.

Variable selection is the classical strategy to discard both redundant and/or uninformative variables. While removing uninformative variables is both theoretically and practically motivated, removing redundant variables makes the selection process instable: the list of selected variables drastically changes with the splitting of the dataset into a training and a test set, leading to a non interpretable classifier. As a solution, an aggregation step can be performed before the selection step. In the aggregation step, variables that share common information are classified into a same cluster and summed up by a summarized variable, and in the selection step informative summarized variables are selected. The difficulty is then to propose an aggregation method that is consistent with the selection method to build an efficient classifier.

We propose a general framework to develop wrapper aggregation methods, where the aggregation specifically depends on the chosen classification algorithm to ensure the coherence between the selection and aggregation steps. We also propose two different algorithms to perform aggregation, for ordered and non-ordered variables respectively. The strategy is illustrated with the CART and kNN algorithms, and we show on several real experiments that the aggregation step leads to more stable and interpretable results.

INTERPRETIVE LATENT SEMANTIC ANALYSIS

Christian MAUCERI
IBM, France
E-mail: mauceri@fr.ibm.com

In this paper we propose to address a recurrent issue of Latent Semantics Analysis (LSA): the difficulty to interpret factors. We use an Interpretive Semantics approach to rethink LSA and propose a new framework for textual semantic analysis taking into account the necessary tradeoffs between human workload and efficiency.

COMPARISON OF PLS LINEAR REGRESSION AND PLS LOGISTIC REGRESSION ON CATEGORICAL DATA

Nicolas MEYER, Myriam MAUMY-BERTRAND
Laboratoire de Biostatistique, Faculty de Medecine, Strasbourg, France
E-mail: Nicolas.Meyer@medecine.u-strasbg.fr

PLS and PLS regression (HÃfÃ¶skuldsson A. Journal of Chemometrics 1988;2 :211-228) are well-known methods used when data are either collinear or when the number of variables is larger than the number of statistical units, as in microarray.

PLS linear regression (PLSLR) is designed to predict a continuous outcome from a set of continuous predictors. Since PLS methods do not rely on any distributional assumptions, PLSLR is sometimes used when both predictors and outcome are categorical. In a purely descriptive goal, this approach is valid. But in a regression setting, it is not clear whether PLS linear regression performances are good, particularly in term of power. Some variants of PLS, such as PLS logistic regression (PLSLogR) (Bastien et al. CSDA 2005;48:17-46) have been designed to explicitly handle categorical only data. Our goal is here to compare the PLS linear regression and the PLSLogR on simulated and classical data sets. The general form of a PLS regression is " $y = XWc$ " where " $T = Wc$ " is a set of components that maximise $cov(y, X)$. The parameters of interest are the " wc " coefficients. Significance of these parameters is computed using permutation methods since the underlying distributions are not known. Simulated data sets are build under (1) null hypothesis of independance and (2) under the alternative hypothesis following a specified pattern between a " y " binomial variable and an " X " predictor matrix of 20 binomial variables. The main results show that, when applied on categorical only data, the PLS logistic regression is slightly more powerfull than the PLS linear regression though the its superiority is not systematic. Particular patterns may appear on distant components. When used as a variable selection methods, PLSLR and PLSLogR are equivalent but the PLSLogR is easier to interpret.

AN EMPIRICAL INVESTIGATION OF THE TRADE-OFF AND PECKING ORDER HYPOTHESES ON ROMANIAN MARKET

Gabriela Maria MIHALCA, Raluca Meda ANTAL
Babes-Bolyai University, Romania
E-mail: gabriela.mihalca@econ.ubbcluj.ro

Starting with the seminal work of Modigliani and Miller (1958, 1963), a growing attention has been given to the manner in which firms adjust their capital structure. When firms adjust their capital structure, they tend to move toward a target debt ratio which is consistent with trade-off theory. According to the trade-off hypothesis profitable firms should have a high level of debt ratio.

Contrary to the trade-off hypothesis, the pecking order theory based on information asymmetry predicts a negative correlation between profitability and leverage.

We test the aforementioned hypotheses on a sample of non-financial Romanian firms, listed on the Bucharest Stock Exchange for 2003 to 2007. Using a panel data analysis, we found that pecking order theory could be successfully applied to the Romanian market. As the pecking order theory predicts, the Romanian profitable firms with a high proportion of tangible assets have a lower debt ratio.

I-POLYA PROCESS AND APPLICATIONS

Leda MINKOVA
Sofia University-Sofia, Bulgaria
E-mail: leda@fmi.uni-sofia.bg

The Inflated-parameter negative binomial process (I-Polya process) as a mixed Polya-Aeppli process is defined. Some basic properties are given. We consider the risk model in which the counting process is the I-Polya process. It is called I-Polya risk model. The joint probability distribution of the time to ruin and the deficit after ruin occurs is studied. The particular case of exponentially distributed claims is given.

INFERENCE IN TWO-SEX BRANCHING MODELS WITH RANDOM CONTROL ON THE NUMBER OF PROGENITOR COUPLES

Manuel MOLINA, Manuel MOTA, Alfonso RAMOS
University of Extremadura, Spain
E-mail: mmolina@unex.es

We introduce a class of two-sex branching processes where a random control on the number of progenitor couples in the population is considered. In addition its theoretical interest, such a class has also several practical implications, especially in population dynamics. For example, in phenomena concerning to inhabit or re-inhabit environments with animal species, it is reasonable to assume that the number of reproductive couples could be affected, in each generation, by some random factors such as weather conditions, food supply, fertility parameters, and so on. First, we determine several relationships among the probability generating functions involved in the mathematical model and then, we provide estimators for the main parameters of the offspring probability distribution. The estimation problem is considered under a parametric point of view, assuming offspring probability distribution belonging to the bivariate power series family. As illustration, some simulated examples are given where the evolution of the estimates is showed.

A NEW ALGORITHM TO SOLVE THE CLASSIFICATION PROBLEMS

Alberto MORENO, Michael SOUZA, Adilson XAVIER
University Rio de Janeiro, Brazil
E-mail: souza.michael@gmail.com

This work considers a new solution methodology for the classification problem associated to the cluster analysis. The classical mathematical formulation for this problem is based on non differentiable optimization problems that are solved by methods that have few precision and efficiency. In this work, by using the hyperbolic smoothing technique, the problem is formulated as completely differentiable problem, whose solution by classic

and more powerful methods can be obtained with great robustness and efficiency. We hope the proposed methodology prove an improvement both reliability and the efficiency in comparison with other alternative procedures.

LOGIC-AND-PROBABILISTIC MODELS OF HETEROGENEOUS MULTIDIMENSIONAL TIME SERIES

Svetlana NEDELKO
Institute of Mathematics, Novosibirsk, Russia
E-mail: nedelko@math.nsc.ru

The method of heterogeneous multidimensional time series adaptive prediction is considered. This method is based on adaptive forming a discrete states set and uses some kind of informativity criterion. The method of adaptation the deciding function to changing of probabilistic properties of the stochastic process is also offered.

EMPIRICAL BOUNDS FOR RISK IN SOME MACHINE LEARNING TASKS

Victor NEDELKO
Institute of Mathematics, Novosibirsk, Russia
E-mail: nedelko@math.nsc.ru

A method of construction empirical bounds for misclassification probability is developed. The method builds a confidence interval using leave-one-out misclassification rate or empirical risk as statistics, no hypotheses on distribution being made. The method was tested on the task of classification via decision trees. Effectiveness of leave-one-out and empirical risk statistics was compared. The approach of empirical bounds on risk was also adopted to some kind of regression task. In opposite to complexity-based estimates by Vapnik and Chervonenkis the method of construction of empirical confidence interval is applicable also to algorithms with infinite capacity.

ON A CLASS OF SDES WITH MARKOVIAN SWITCHING AND APPLICATIONS TO ELECTRONICS CIRCUITS

Romeo NEGREA, Ciprian PREDA, Horia CARSTEA, Ciprian HEDREA
Department of Mathematics, University Politehnica Timisoara, Romania
E-mail: negrea@math.uvt.ro

Noise in dynamical system is usually considered a nuisance. However, in certain nonlinear systems, including electronic circuits and biological sensory systems, the presence of noise can enhance the detection of weak signals. The phenomenon is termed stochastic resonance and is of great interest for electronic instrumentation.

The essential ingredient for the stochastic resonance is a nonlinear dynamical system, which typically has a period signal and noise at the input and output that is a function of the input as well as the internal dynamics of the system. The nonlinear component of the dynamical system is sometimes provided by a threshold which must be crossed for the output to be changed or detected. A nonlinear system is essential for stochastic resonance to exist, since in a system that is well characterized by linear response theory, the signal-to-noise ratio at the output must be proportional to the signal-to-noise ratio at the input.

In this paper, we study a nonlinear system under regime switching and subject to an environmental noise. We will consider some more general conditions for the coefficient functions and prove a result on the existence and the pathwise uniqueness and some stability properties extended some similarly results on linear systems. Using the benefit of stochastic resonance we apply the obtained results to control an electronic circuits (by Chua's type) with different amplifiers.

CHANGE POINT DETECTION BY SPARSE PARAMETER ESTIMATION

Jiri NEUBAUER, Vitezslav VESELY
University of Defence, Czech Republic
E-mail: neubauer@email.cz

The contribution is focused on change point detection in one-dimensional stochastic processes by sparse parameter estimation in overcomplete models. Stochastic processes with changes in the mean are estimated by Heaviside functions. The Basis Pursuit algorithm is used to get sparse parameter estimates. The mentioned method of change point detection in stochastic processes is compared with standard methods by simulations.

BATCH PROCESS MONITORING BY THREE-WAY DATA ANALYSIS APPROACH

Ndeye NIANG, Flavio FOGLIATTO, Gilbert SAPORTA
CNAM, France
E-mail: ndeye.niang_keita@cnam.fr

We propose a non parametric quality control strategy based on the three way method STATIS and convex hull peeling for monitoring batch processes with constant or variable duration. The method is illustrated on a simulated data set and on a real one.

SELECTION OF THE LEARNING DATABASE ELEMENTS FOR EL-DORRA NEURAL NETWORK TECHNOLOGY GLOBAL STATIONARITY AND EXISTENCE OF THRESHOLD ARMA MODELS

Marcella NIGLIO, Cosimo Damiano VITALE
DISES, University of Salerno, Italy
E-mail: mniglio@unisa.it

The threshold ARMA (TARMA) model, firstly introduced in Tong (1983), can be considered as generalization, in nonlinear domain, of the linear ARMA model widely discussed in Box and Jenkins (1976).

Despite the interest raised by this class of models, the conditions for their stationarity have been investigated only later in Brockwell et al. (1992), Liu and Susko (1992) and Ling (1999) whereas, more recently Amendola et al. (2008), have weakened these too strong conditions. The most part of results given in the cited literature, connect the global stationarity of the process to the stationarity of each single regime. In this domain, an interesting issue of the TARMA model, that can be empirically appreciated through simulations, is that the global stationarity of the process can be even reached in presence of regimes that are locally nonstationary. In our knowledge, this problem has been faced in literature considering threshold autoregressive (TAR) processes, having unit roots in their regimes, for which proper tests are proposed to detect the (global) stationarity (among the others see Balke and Fomby (1997), the seminal paper of Caner and Hansen (2001) recently revisited in Pitarakis (2008), Kapetanios and Shin (2006), Bec et al. (2008)). In this context we provide the conditions, on the model parameters, under which the global (but not local) stationarity of the TARMA process is obtained. We further focus the attention on the existence of the TARMA process: it is based on the consideration that each regime needs to have a sufficient number of observations that warrant not only the use of a nonlinear TARMA structure (instead of the simpler linear ARMA model) but even the estimation of its parameters.

References

1. Amendola, A. Niglio, M, Vitale, C. (2008). Statistical properties of threshold models. *Communications in Statistics: Theory and Methods*, accepted for publication.
2. Bec, F. Guay, A. Guerre, E. (2008). Adaptive consistent unit-root tests based on autoregressive threshold model. *Journal of Econometrics*, 142, 94-133.
3. Balke, N.S. Fomby, T.B. (1997). Threshold cointegration. *International Economic Review*, 38, 627-645.
4. Box, G.E.P. Jenkins, G.M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden-Day.
5. Brockwell, P.J. Liu, J. Tweedie, R.L. (1992). On the existence of stationary threshold AR-MA process. *Journal of the Time Series Analysis*, 13, 95-107.
6. Caner, M Hansen, B.E. (2001). Threshold autoregression with a unit root. *Econometrica*, 69, 1555-1596.
7. Kapetanios, G. Shin, Y. (2006). Unit root tests in three-regime SETAR models. *The Econometrics Journal* 9, 252-278.
8. Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model. *Journal of Applied Probability*, 36, 688-705.

9. Liu, J. Susko, E. (1992). On strict stationarity and ergodicity of a non-linear ARMA model. *Journal of Applied Probability*, 29, 363--373.
10. Pitakaris, J.Y (2008). Comment on: threshold autoregression with a unit root. *Econometrica*, 76, 1207-1217.
11. Tong, H. (1983). *Threshold models in nonlinear time series analysis*. London: Springer-Verlag.

ROUGH FUNCTIONS AND p-VARIATION CALCULUS

Rimas Norvaisa
Institute of Mathematics and Informatics
E-mail: norvaisa@ktl.mii.lt

For any positive number p , the p -variation of a real-valued function f defined on the interval $[0,1]$ is the same as the total variation except that one sums the p -th powers of increments of f corresponding to partitions of $[0,1]$. For example, if $p > 2$ then almost every sample function of a Brownian motion has bounded p -variation and its 2-variation is infinite. The aim of p -variation calculus is to find how smooth are various non-linear operators acting between Banach spaces of functions of bounded p -variation. One such example is the operator induced by the product integral of a function, which gives a solution to a linear integral equation. This operator is analytic when acts from the space of functions of bounded p -variation to itself and $p > 2$. The p -variation calculus can be used to extend the theory of statistical functionals (also called the delta-method), usually treated with respect to the uniform norm. We show that many useful statistical functionals become smooth when treated with respect to the p -variation norm. Also we show that the Donsker invariance principle extends to the p -variation norm.

FUNCTIONAL MODELLING IN ENVIRONMETRICS

Francisco A. OCANA-LARA, Mariano J. VALDERRAMA,
Francisco M. OCANA-PEINADO, Manuel ESCABIAS
University of Granada, Spain
E-mail: focana@ugr.es

An usual approach for functional regression of a stochastic process $Y(t)$ versus a vector variable $(X_1(t), X_2(t), \dots, X_n(t))$ consists of developing a multi-step marginal procedure for explaining $Y(t)$ in terms of a certain $X_i(t)$ and then the procedure is step-wise applied to the residuals. The relation among the coefficients of linear determination is derived and the equivalence between a functional linear regression and a principal component prediction model is also proved. Moreover, we estimate a functional model for pollen concentration in a certain interval of time from its past history as well as the environment temperature.

IDENTIFICATION OF WINTER PRECIPITATION PATTERNS WITH A FACTOR ANALYSIS (FA) IN A MOUNTAIN AREA OF SIERRA NEVADA (SOUTHERN IBERIAN PENINSULA)

Marc OLIVA, Paulo PEREIRA
Landscape Research and Mediterranean Mountain Paleoenvironment Research Group, Spain
E-mail: pereiraub@gmail.com

Our research is focused in Sierra Nevada, the range which holds the highest summits in southwestern Europe, with peaks exceeding 3000 m at 37 latitude. Sierra Nevada plays a decisive role in the climate of the southern Iberian Peninsula; the singular geographical characteristics of this massif determine a wide range of microclimates in the surrounding area. The main interest in studying the factors involved in precipitation in this region lies in its geographical position in an area with crossing influences: between Europe and Africa, between the Atlantic Ocean and the Mediterranean Sea, between the subtropical high-pressure belt and the mid-latitude westerlies. To deal with such a huge dataset, we use a Factor Analysis (FA) which is an excellent tool to reduce the number of potential involved variables and to detect the relationships between

them. To obtain a winter precipitation pattern (December, January and February) in this mountain area, we applied a FA with the extraction method of Principal Components (PCA) in order to identify a possible regionalization model. We analyze up to 23 climate series geographically distributed between 1961 and 2005 considering the total winter precipitation. The FA extracted two PCA: the first of them explained 75.22% of the variance and the second one the rest 24.78%. We tested several rotation methods (VARIMAX, BIQUERTIMAX, QUARTIMAX and EQUAMAX) to identify a clear behavior of the loadings (factors with higher loadings for some variables and low for another's). Among all these methods, the VARIMAX provided the best results. Only loadings > 0.7 were considered for the analysis of each factor.

Our results showed two clear teleconnection patterns that influence the winter precipitation in Sierra Nevada area. The first factor presented higher positive loadings in the western stations and the second one in the eastern part of the massif. We observed that the first factor explains very well the influence area of the North Atlantic Oscillation (NAO) in Sierra Nevada, pattern that determines the winter rainfall in the Northern Hemisphere. The second factor is associated with the Western Mediterranean Oscillation (WeMO), a pattern that rules the precipitation in the western part of Mediterranean sea. After the identification of these patterns we analyzed the winter rainfall trend in both regions. A negative trend is detected since 1951 mainly in the eastern stations, which means that the NAO and especially the WeMO influence in southern Iberian Peninsula has decreased in the second half of the XX century.

FITTING THE ACADEMIC PASS RATES BY AN EXTENSION OF THE BIVARIATE BETA-BINOMIAL DISTRIBUTION

Maria Jose OLMO-JIMENEZ, Ana Maria MARTINEZ-RODRIGUEZ,
Antonio CONDE-SANCHEZ, Jose RODRIGUEZ-AVI,
Silverio VOLCHEZ-LOPEZ

Department of Statistics and Operations Research. University of Jaen, Spain
E-mail: mjolmo@ujaen.es

One of the main indicators of the academic achievement of the students is the pass rate. In this work we use the bivariate Beta-Binomial distribution (BivBB) proposed by Alanko and Lemmens (1996) to model the number of subjects passed for every student in the University of Jaen (Spain) during two consecutive academic years (2004/2005 and 2005/2006). This distribution has linear regression curves and is symmetric when the range of the two marginal variables is the same. These two properties may be a disadvantage for fitting a data set, so our purpose is to generalize this distribution so that it is not so restrictive. Thus, the proposed distribution is also generated by Appell's hypergeometric function F_1 , is not symmetric and is more flexible than the BivBB. Moreover, it can be obtained as a Binomial mixture, so we may establish the sources of variability of the data modeled.

MIGRATIONS OF HETEROGENEOUS POPULATION OF DRIVERS ACROSS CLASSES OF THE BONUS-MALUS SYSTEM

Wojciech OTTO
University of Warsaw, Poland
E-mail: wotto@wne.uw.edu.pl

Traditionally, migrations of a single driver across Bonus-Malus classes are modeled as a simple homogeneous Markov Chain. Generalization to the case when the modeled object is a population of drivers, where all drivers are characterized by the same frequency of claims, is very straightforward. In practice however, drivers differ. Moreover, difference in frequencies characterizing individual drivers are the main reason of BM systems applied in motor insurance.

The presentation is devoted to report results of the study focused on consequences of population heterogeneity. At first the probabilistic model adequate in the case of heterogeneity is introduced. Consequences of allowing for heterogeneity are presented then by showing how much properties of the migration process depart from properties of the homogeneous Markov chain adequate in the case when population is homogeneous. Finally, some consequences of heterogeneity that concern estimators of transition probabilities from empirical data are presented.

CLASSIFICATION AND CLUSTERING OF GARCH TIME SERIES

Demetrios PAPANASTASSIOU
University of Macedonia, Greece
E-mail: papanast@uom.gr

There is practical interest to understand and measure the risk involved in financial time series. Risk is usually related to the series volatility, and, in this respect, the most widespread tool to represent risk is the family of GARCH models initiated by Engle (1982) and Bollerslev (1986). We first note that the parameters of a simple GARCH(1,1) model entail information on several other than simple conditional variance risk related aspects, such as kurtosis, volatility persistence and extremal behaviour. This note supports the idea to classify and group financial series according to their proximity to certain meaningful GARCH parameter values. To this end, in this paper we introduce a distance measure between two different Gaussian GARCH parameterizations based on the Kullback-Leibler information measure. A series is classified into one of several predefined groups according to the distance of the fitted GARCH from these groups. We give an asymptotic expression for the error rate, and a simulation study explores the efficiency of the discriminant rule. We find that the empirical error rate from the simulation repetitions decreases with sample size and it is relatively close to the anticipated theoretical. The accuracy of the latter improves by a simple bootstrap correction. We also test the robustness of the rule with data generated by non-Gaussian processes. A symmetric version of our distance is used for clustering the series into homogenous groups, and, again, a simulation study shows that our procedure is reliable. We conclude with a real data example.

EVALUATION AND COMPARISON OF SOME INTERPOLATION METHODS TO PREDICT PH AND ELECTRICAL CONDUCTIVITY SPATIAL DISTRIBUTION RELEASED BY ASHES AFTER A WILDFIRE

Paulo PEREIRA, Xavier OBEDA
Mediterranean Environmental Research Group,
E-mail: pereiraub@gmail.com

Interpolation methods are very important in order to predict data in places where we do not have samples. The great majority of the studies use the Geostatistical methods to data interpolation because they identify a spatial variation of the variable. However, these methods are not always the best predictor for unsampled points. The aim of this study is test the “best method“ of interpolation testing some deterministic methods - Local Polynomial (LP), Inverse Distance to a Power (IDP), Nearest Neighbour (NN) and Polynomial Regression (PR) - and a Geostatistical method - Ordinary Kriging (OK) - of the pH and Electrical Conductivity (EC), released from ash collected after a wildfire in Portugal. We designed a plot of 27 x 9 m in the burned area and collected a total of 40 samples to test our study. After laboratory analysis, data Gaussian distribution observed, considered normal at a $p < 0.05$, according to the Shapiro-Wilk test. Since, the data of pH and EC were normally distributed, no transformation was applied and we use the original data to test the interpolation methods. The performance of each interpolation method is evaluated comparing the deviation of the estimates from the measure data (errors) applying a Cross validation procedure. Hence, from the estimated errors we applied some accuracy tests as the Mean Error (ME) - which determine the degree of bias in the estimates -, the Root Mean Square Error (RMSE) - that measure the size of the error and it is sensitive to outliers influence. The best interpolation method is the one with RMSE closer of 0 - and the Relative Improvement (RI) - that compare the best method observed with the others. The results showed that in general for pH, the ME was lower in all methods, reaching to a maximum in NN (0.0065) and lower in PR (0.0006). RMSE is lower in LP (0.124) and higher in NN (0.207), which means that the first method is the “best predictor“ than the others. Observing the RI, the worst interpolation method is the NN (66.22%) and the closest is OK (5.40%). In relation to EC, we observed also that data have low ME, with a maximum of in NN (0.215) and a minimum in PR (0.0000). In relation to RMSE, the smaller value was observed in PR (0.467) and the higher in NN (0.863). Hence, for EC, the best predictor method is PR and the worst is NN with a RI of 84.82%. The closest predictor method of PR is IDP (9.24%). Overall, we observed that the polynomial interpolation methods were the best predictors and the worst in both variables was NN. Due to our results in our opinion the geostatistical methods are not always the best data

interpolators. Interpolation accuracy tests are important to define the best interpolation method, before analyse spacial data distribution.

SHOCK AND WEAR MODELS UNDER MATRIX-ANALYTIC METHODS

Rafael PEREZ-OCÓN, M. Carmen SEGOVIA
University of Granada, Spain
E-mail: msegovia@ugr.es

A device subject to shocks arriving randomly is considered. Every shock can be fatal or not. The shocks follow a Markovian arrival process. When the shock is fatal, the device is instantaneously replaced. The Markov process governing the shocks is constructed, and the stationary probability vector calculated. The probability of the number of replacements during a time is determined. Particular cases of interest are studied and a numerical application is performed. The expressions are given in algorithmic form. Computational aspects are introduced. This model extends other previously considered in the literature.

CATASTROPHIC RISK ESTIMATION USING SKEW t -COPULA

Gaida PETTERE, Tonu KOLLO
Department of Engineering Mathematics, Riga Technical University, Latvia
E-mail: gaida@latnet.lv

Following CEIOPS QIS4 technical specifications Catastrophic risk in nonlife insurance is defined as a risk which stems from extreme or irregular events that are not sufficiently captured by the charges for premium or reserve risk. In the same document three methods are described to define capital charge for catastrophic risk. Standard method implies capital calculations using premiums written in each business line and special coefficients equal for all companies throughout all Europe. But business characteristics for each company are different. Our aim is to take into account portfolio structure of a company and to show under the internal model how necessary capital can be modeled using skew t -copula for an insurance portfolio.

EXACT SMOOTHING IN HIDDEN CONDITIONALLY MARKOV SWITCHING CHAINS

Wojciech PIECZYNSKI
Institut Telecom, Telecom SudParis, 8, rue Charles Fourier, 91 000 Evry, France
E-mail: Wojciech.Pieczynski@it-sudparis.eu

The problem considered in this paper is the problem of the exact calculation of smoothing in hidden switching state-space systems. There is a hidden state-space chain X , the switching Markov chain R , and the observed chain Y . In the classical, widely used “conditionally Gaussian state-space linear model (CGSSLM) the exact calculation with complexity linear in time is not feasible and different approximations have to be made. Different alternative models, in which the exact calculations are feasible, have been re-cently proposed (2008). The core difference between these models and the classical ones is that the couple (R, Y) is a Markov one in the recent models, while it is not in the clas-sical ones. Here we propose a further extension of these recent models by relaxing the hypothesis of the Markovianity of X conditionally on (R, Y) . In fact, in all classical mod-els and as well as in the recent ones, the hidden chain X is always a Markov one conditionally on (R, Y) . In the proposed model it can be of any form. In particular, different “long memory“ processes can be considered. In spite of this larger generality, we show that the smoothing formulae are still calculable exactly with the complexity polynomial in time.

EXACT FILTERING AND SMOOTHING IN MARKOV SWITCHING SYSTEMS HIDDEN WITH GAUSSIAN LONG MEMORY NOISE

Wojciech PIECZYNSKI, Noufel ABBASSI, Mohamed Ben MABROUK
Telecom SudParis, France
E-mail: Wojciech.Pieczynski@it-sudparis.eu

Let X be a hidden real stochastic process, R be a discrete finite Markov chain, Y be an observed chain. The problem of filtering and smoothing is the problem of recovering both R and X from Y . In the classical models the exact computing with linear - or even polynomial - complexity in time index is not feasible and different approximations are used. Different alternative models, in which the exact calculations are feasible, have been recently proposed (2008). The core difference between these models and the classical ones is that the couple (R, Y) is a Markov one in the recent models, while it is not in all the classical ones. Here we propose a further extension of these models. The core point of the extension is the fact that the observed chain Y is not necessarily Markovian conditionally on (X, R) and, in particular, the long-memory distributions can be considered. We show that both filtering and smoothing are computable with complexity polynomial in the number of observations in the new model.

FINANCIAL VALUATION OF GUARANTEED LIFELONG WITHDRAWAL BENEFIT OPTION

Gabriella PISCOPO
Universite degli studi di Napoli Federico II
E-mail: gabriella.piscopo@unina.it

In the light of the growing importance of the Variable Annuity (VA) market, the aim of this paper is to define a fair price for a Guaranteed Lifelong Withdrawal Benefit (GLWB) option embedded in a VA. The VA, whose benefits are based on the performance of a underlying fund, are very attractive, because they provide a participation in the stock market and also a partial protection against the downside movements of the interest rates or the equity market. The GLWB offers a lifelong withdrawal guarantee; therefore, there is no limit for the total amount that is withdrawn over the term of the policy, because if the account value becomes zero while the insured is still alive he can continue to withdraw the guaranteed amount annually until death.

In order to price the GLWB option I use the standard No-arbitrage models of mathematical finance, in line with the tradition of Boyle and Schwartz (1997) that extend the Black-Scholes framework to insurance contract. The main difference is that for the option embedded in VA products the fee is deducted ongoing as fraction of asset, instead in the Black and Scholes approach the premium is paid up-front. The approach follows the recent actuarial literature on the valuation of VA products: Bauer et al. (2006); Chen et al. (2008), Coleman et al. (2006); Holz (2006), Milevsky and Posner (2001), Milevsky M.A and Promislow S.D (2001), Milevsky and Salisbury(2002). In this regard, Milevsky and Salisbury (2006) decompose the GMWB option into a Quanto Asian put plus a generic term-certain annuity in order to calculate its fair price. Furthermore, they find that the Guaranteed Withdrawal Minimum Benefit options are usually underpriced in the market, in contrast to the general belief that the guarantees embedded in VA contracts are all overpriced (Clements (2004)). In order to price options embedded in Variable Annuity contracts many authors use numerical PDE methods (see Dai (2008), Chen et al. (2008), Milevsky and Posner (1998), Milevsky and Salisbury (2006), Nielsen and Sandmann (2003)), others exploit Monte Carlo simulations (Milevsky and Panyagometh (2001), Milevsky and Posner (2000)). I choose to follow the latter approach because we don't give attention to the numerical algorithm but we fix our analysis on the financial and actuarial nature of the problem. Therefore, we develop two formulas in order to price the GLWB option, both within the framework of No Arbitrage pricing. First, I take a static approach that assumes policyholders take a static strategy, i.e. the withdrawal amount is always equal to the guaranteed amount. In this case we show the product can be decomposed into a life annuity plus a portfolio of Quanto Asian Put Options, with decreasing strikes and increasing expiration dates. The opposite assumption is that all investors are rational and maximize the embedded option value by surrendering the product at an optimal time, when the surrender value exceeds the value of future benefits. In both cases I calculate the fair insurance fee making equal the

initial price of the product to the zero value of the future cash flows. As Milevsky and Salisbury note, the real fee has to lie between the static and dynamic embedded option cost; in fact, the policyholder can behave with a high level of irrational lapsation.

The main contribution of this work lies in bifurcating the product into the life annuity component and the derivatives components and calculate the fair insurance fee. The conclusion is that the GLWB issued on the USA market are underpriced and this appears regardless of whether we take a static or dynamic approach. For example, the numerical results show that the No Arbitrage cost of a GLWB issued to a policyholder aged 60 would range between 79 and 145 basis points assuming a sub-account volatility in line with the average of the sub-account volatility for the universe of variable annuity products, while most product in the USA market only charge 50-70 basis points. I compute the probability of ruin, i.e. the probability that the fund hits zero and the insurer has to pay the amount guaranteed, if the insurance fee is that normally charged by the market. This probability oscillates between 48% and 57% under different expected investment returns and increases with the sub-account volatility; given the long-dated nature of the embedded option it is likely that the volatility would increase and thus the risk for the insurer to pay the guaranteed amount is very high. This results indicate that the market fees are not sufficient to cover the market hedging cost of the guarantee. Of course, the pricing model does not allow for more sophisticated financial hypothesis, such as stochastic volatility or jumps in the fund process and term-structure effects, but as Milevsky and Salisbury (2006) I am confident that these considerations will only increases the price of the embedded option. The same effect would be obtained with the introduction of an actuarial model allowing for the longevity risk. So, I conclude by arguing that the current price of GLWB is not sustainable for insurers and the fees have to increase in order to avoid arbitrage opportunities.

References

1. Bauer D., Kling A., Russ J., 2006. A universal pricing framework for guaranteed minimum benefit in variable annuities. Working Paper, Ulm University.
2. Boyle P.P, Schwartz E., 1977. Equilibrium prices of guarantees under equity-linked contracts. *Journal of Risk and Insurance* 44 (2), 639-680
3. Chen Z., Forsyth P., 2007. A numerical scheme for the impulse control formulation for pricing variable annuities with a Guaranteed Minimum Withdrawal Benefit. *Numerische Mathematik* (in press)
4. Chen Z., Vetzal K., Forsyth P.A., 2008. The effect of modelling parameters on the value of GMWB guaranteed. *Insurance: Mathematics and Economics* 43, 165-173.
5. Clements J., 2004. For a conservative investment, variable annuities are too costly. *Wall Street Journal* C1, January 21
6. Coleman T.F., Li Y., Patron M., 2006. Hedging guarantees in variable annuities under both equity and interest rate risks. *Insurance: Mathematics and Economics* 38, 215-228
7. Cramer E., Matson P., Rubin L., 2007. Common practices relating to FASB statement 133, Accounting for Derivative Instruments and Hedging Activities as it Relates to Variable Annuities with Guaranteed Benefits. Practice Note, American Academy of Actuaries.
8. Dai M., Kwok Y.K., Zong J., 2008. Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance* (in press).
9. Gerber H.U., Shiu E., 2003. Pricing dynamic investment fund protection. *North American Actuarial Journal* 4 (2), 60-77
10. Hanif F., Finkelstein G., Corrigan J., Hocking J., 2007. Life Insurance, global variable annuities. Published by Morgan Stanley Research in co-operation with Milliman.
11. Ho T., Lee S.B., Choi Y., 2005. Practical consideration in managing variable annuities. Working paper, Thomas Ho Company.
12. Holz D., Kling A., Rub J., 2006. GMWB For Life. An Analysis of Lifelong Withdrawal Guarantees. Working Paper ULM University
13. Ledlie M.C., Corry D.P., Finkelstein G.S., Ritchie A.J., Su K., Wilson D.C.E, 2007. Variable Annuities. Presented to the Institute of Actuaries , London
14. Milevsky M., Posner S.E., 2001. The Titanic Option: Valuation of the Guaranteed Minimum Death Benefit in Variable Annuities and Mutual Funds. *The journal of Risk and Insurance* 68, 1, 91-126.
15. Milevsky M., Salisbury T.S., 2002, The Real Option to Lapse and the Valuation of Death-Protected Investments. Working Paper York University and the Fields Institute, Toronto.

16. Milevsky M.A., Panyagometh K., 2001. Variable Annuities versus Mutual Funds: a Monte Carlo Analysis of the Option.
17. Milevsky M.A., Posner S., 1998. Asian option, the sum of lognormals and the reciprocal gamma distribution. *Journal of Financial and Quantitative Analysis* 33 (3), 409-422.
18. Milevsky M.A., Posner S., 2000. *The probability of Fortune: Financial Strategies with the Best Odds*. Stoddart Publishing Co., New York.
19. Milevsky M.A., Promislow S.D., 2001, Mortality derivatives and the option to annuities. *Insurance: Mathematics and Economics* 29, 299-318
20. Milevsky M.A., Salisbury 2006. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics* 38, 21-38
21. Nielsen J.A., Sandmann K., 2003. Pricing bounds on Asian options. *Journal of Financial and Quantitative Analysis* 38 (2), 449-473
22. Sloane W.R., 1970. Life Insurers, Variable Annuities and Mutual Fund: a Critical Study. *Journal of Risk and Insurance* 37, S.99
23. Wang D., 2008. Pricing Variable Annuity with Embedded Guarantees- a Case Study. Milliman Research.
24. Windcliff H., Forsyth P.A., Vetzal K.R., 2001. Valuation of segregated funds: shout options with maturity extensions. *Insurance: Mathematics and Economics* 29 (2), 1-21

REGRESSION MODELS FOR CATEGORICAL FUNCTIONAL DATA

Cristian PREDA
University of Lille 1, France
E-mail: cristian.preda@polytech-lille.fr

Functional data represented by observations of a categorical random variable X over some interval of time $[0, T]$ is considered. Elements derived from the transition probability function of the underlying stochastic process allow us to extend some tools from finite multivariate analysis such as factorial analysis and regression. The case of stationary reversible Markov continuous-time processes with finite space state is detailed.

THE WEIBULL-LOGARITHMIC DISTRIBUTION IN LIFETIME ANALYSIS AND ITS PROPERTIES

Vasile PREDA, Roxana CIUMARA
University of Bucharest, Romania
E-mail: roxana_ciumara@yahoo.com

In this paper we propose a new two-parameter lifetime distribution, namely Weibull-Logarithmic distribution. The study of lifetime and residual lifetime distributions play a central role in life insurance mathematics, and moreover they could be applied in the biological and engineering sciences. Along the years, there have been developed many parametric models for (residual) lifetime distributions and they were analyzed with respect to different characteristics. Among the papers revealing interesting properties of several (residual) lifetime distributions, we underline the work of Lomax (1954), Marshall and Proschan (1965), Saunders and Myhre (1983), Ebrahimi (1996), Adamidis and Lukas (1998), Kus (2007), Tahmasbi and Rezaei (2008).

Our new lifetime distribution is obtained by mixing Weibull and logarithmic distributions. Then we find the properties of this distribution, by deriving the cumulative distribution function, probability density function, the median, the moment generating function and, consequently, the k -th moments. We give a special attention to the survival, hazard and mean residual functions. All the properties obtained for Weibull - logarithmic distribution are compared to the ones of a simple Weibull distribution and we emphasize those situations where Weibull-logarithmic model is more appropriate to be used. Also we propose methods for estimating the parameters of the Weibull-logarithmic distribution and we derive their features.

Furthermore, taking into account that an important component factor of a lifetime distribution is the hazard rate function, we compare this measure to an alternative one - proposed in the paper of Ebrahimi (1996) - that is, the conditional Shannon's measure of uncertainty.

In the last part of the paper, we perform some simulation studies and we comment on the results. Guidelines for further research are also pointed out at the end of the paper.

AN UNIFIED APPROACH FOR OPTIMAL RETENTION IN A STOP-LOSS REINSURANCE UNDER THE VAR AND CTE RISK MEASURES

Vasile PREDA, Silvia DEDU, Roxana CIUMARA
University of Bucharest, Romania
E-mail: vasilepreda0@gmail.com

The aim of this paper consists in solving the problem of deriving the optimal retention in stop-loss reinsurance. We develop a combined optimization criterion, based on minimizing a linear convex combination of VaR and CTE risk measures. Necessary and sufficient conditions for the existence of the optimal solution are established. This approach allows us to obtain different optimal solutions, depending on the proportion the two risk measures are used. The solution obtained extends the results of Cai and Tan (2007).

Furthermore, a modified risk measure is defined. The problem of deriving the optimal retention under this risk measure is considered. The results obtained are illustrated using simulations. Computational results are provided.

PROPOSAL SELECTION IN MCMC SIMULATION

Vilmos PROKAJ
Eotvos University Budapest, Hungary
E-mail: prokaj@cs.elte.hu

A simple application of the stochastic approximation algorithm is presented for tuning the proposal in MCMC simulation. In this note we investigate the problem of tuning the free parameters in MCMC simulation. We focus on only the probably most frequently used random walk Metropolis-Hastings (RWMH) algorithm. First, we briefly recall this algorithm. Then we propose a stochastic approximation algorithm for self tuning the proposal's spread parameter, and we adopt this method for tuning the proposal's covariance in case of block update. Finally an example is presented. The proposed algorithm is not an adaptive MCMC, as in our case the chain is time-variant only in the 'burn-in' period, and our aim is to find the best fitting RWMH sampler for the given problem. Nevertheless the ideas we use are similar to those available in the literature, see eg. Haario et al. (2001), Andrieu and Atchade (2007) and Atchade and Rosenthal (2005).

ON THE STUDY OF THE NUMBER OF OCCURENCES OF THE LOCAL OPTIMUM TO SPOT GENETIC ALGORITHM CONVERGENCE

Christelle REYNES, Robert SABATIER
MTi, Unite Inserm, Paris 7 Diderot U973, France
E-mail: christelle.reynes@univ-paris-diderot.fr

Genetic Algorithm convergence remains a major issue in any application of this optimization method. The most usual method is to empirically determine a maximum number of generations, a maximum computation time or to determine a threshold of non significance of fitness value evolution. Each of these criteria has to be reset for each new application leading to increase the analysis time which is generally a problem in genetic algorithms context.

The goal of this paper is to design a new stopping criterion based on the modelling of elitist genetic algorithms (the best solution is compulsory part of the next generation) by the means of Markov chains.

Studying genetic algorithms in a markovian context is quite common but the proposed criterion will use it in a new way.

This criterion studies the distribution of the number of occurrences of the locally best solution. Indeed, when the algorithm is far from convergence, the local optimum frequently changes and it does not have enough time to “invade“ the population. On the opposite, the nearer the local optimum to the global one, the longer it stays in the population. In such a case, thanks to selection and elitism, more and more copies of the local optimum are present in the population.

In order to use this observation, the number of occurrences of the locally best solution is modeled under the assumption of non-convergence. The algorithm stops when the probability of the observed value under this assumption is too small.

It is important to notice that this criterion does not guarantee that the optimal solution has been found. The aim is to make a balance between theoretical results (most of time impossible to use in practical cases) and usual practical criteria (without any theoretical basement). It is obvious that for real optimization issues, it is impossible to be sure to reach the global optimum with heuristic methods but it is of big interest to have criteria to ensure a good quality of the final solution. Simulations have been performed to check the efficiency and accuracy of this new criterion and really encouraging results have been obtained. For example, two ways of optimizing the same function (leading to very different convergence speeds) did not make the criterion fail.

COUNT DATA REGRESSION MODELS IN THE SPANISH BANK SYSTEM. APPLICATION FOR THE NUMBER OF OFFICES PER MUNICIPALITY IN ANDALUSIA (SPAIN)

Jose RODRIGUEZ-AVI, Maria Jose OLMO-JIMENEZ,
Ana Maria MARTINEZ-RODRIGUEZ, Antonio CONDE-SANCHEZ,
Antonio Jose SAEZ-Castillo
University of Jaen, Spain
E-mail: jravi@ujaen.es

The Spanish bank system is mainly composed of two kinds of institutions, one of them private – the banks- and other public –the savings banks-. In this work we focus on the study of the count data variables “number of bank offices per municipality” and “number of savings bank offices per municipality. We propose several regression models for discrete variables, such as Poisson, Quasi-Poisson, or Negative binomial in order to explain these variables in terms of some other economic variables (market share, economic activity, industrial rate, tourist rate, etc.), and we try to interpret the results obtained. Specifically, we analyze data referred to the autonomous region of Andalusia (Spain) in 2007.

MEASURING OUTPUT GAP IN LITHUANIA 1997 - 2007

Jurga RUKSENAITĖ
Vilnius Gediminas Technical University, Lithuania
E-mail: jurga.ruksenaite@fm.vgtu.lt

The output gap is well-known theoretical concept which plays an important role in many macro-models related to fiscal and monetary evaluation of country’s economy. When the economy is running an output gap, either positive or negative, it is thought to be running at an inefficient rate as the economy is either overworking or underworking its resources. Economic theory suggests that positive output gap will lead to inflation as production and labour costs rise. The OBJECT of the research is the measurement of Lithuania’s output gap by different statistical methods: univariate (time trend, Hodrick-Prescott filter and unobserved component model) and multivariate (production function). Results are compared identifying differences and similarities between mentioned models.

EMPIRICAL BAYES APPROACH FOR SMALL RATES ESTIMATION

Leonidas SAKALAUSKAS
Institute of Mathematics & Informatics, Lithuania
E-mail: sakal@ktl.mii.lt

The paper considers numerical issues of the empirical Bayesian approach model, applied in the estimation of small rates. The condition for non-singularity of Bayesian estimates is derived and an efficient iterative algorithm for estimation is described. The clustering algorithm is also developed, using the property of the Poisson-Gaussian model to treat probabilities of events in populations to be the same, if the variance of rates in populations is small. The approach considered is illustrated by an application to the analysis of homicides and suicides data in Lithuania in 2003-2004.

ESTIMATION OF SKEW-t DISTRIBUTION BY MONTE-CARLO MARKOV CHAIN APPROACH

Leonidas SAKALAUSKAS, INGRIDA VAICIULYTE
Institute of Mathematics & Informatics, Lithuania
E-mail: sakal@ktl.mii.lt

The algorithm for application of Monte-Carlo Markov Chain (MCMC) to estimation of parameters of skew t-distribution is developed. In general, the skew-t distribution is represented by multivariate skew-normal distribution with covariance matrix depending on parameter distributed according to inverse-gamma distribution (Azzalini (1985)). Using this representation, the density of skew t-distribution is expressed through multivariate integral which is convenient to estimate numerically by Monte-Carlo simulation. Thus, the likelihood function, depending on some data sample, might be estimated by Monte-Carlo method, too. Next, the MCMC procedure is developed for recurrent estimation of skew t-distribution, where the Monte-Carlo sample size is regulated so that to ensure the convergence and to decrease the total amount of Monte-Carlo trials. Besides, the rule for termination of the chain is proposed testing the statistical hypotheses about insignificant change of estimates of parameters in two steps of the algorithm.

ADEQUACY TESTING OF TWO METHODS FOR MODELING SPATIAL TIME SERIES

Laura SALTYTE
Klaipeda University, Lithuania
E-mail: laura@ik.ku.lt

Referring to earlier suggested modeling technique named spatial connection, which can be applied to univariate time series data observed at different spatial locations (spatial time series), research of adequacy is presented. Research is based on RMSE results that are calculated for few different predictions.

SUPERVISED AND UNSUPERVISED LINEAR METHODS FOR FUNCTIONAL DATA

Gilbert Saporta
Conservatoire National des Arts et Métiers, France
E-mail: gilbert.saporta@cnam.fr

Functional data occurs when we observe curves or paths from a stochastic process X_t . We will assume that t takes continuously its values in an interval $[0;T]$. Karhunen-Loeve decomposition provides then an unsupervised method for describing this kind of data and is the continuous counterpart of principal component analysis. If for each curve or path we have in addition a single response variable Y , we have a regression problem when Y is numerical, a supervised classification problem when Y is categorical. Linear methods looks for predictors which may be expressed as an integral sum $\int_0^T \beta(t)X_t dt$. The problem is not

new and comes back to Fisher (1924). The strong autocorrelations between predictors leads to inconsistent estimation of the parameters. Since the works of Ramsay & Silverman (1997), many techniques have been applied to solve these kind of problems, mostly by using explicit regularization techniques. We will focus here on linear methods based on an orthogonal decomposition of the predictors. The use of components derived from the Karhunen-Loeve decomposition is, for functional data, the equivalent of principal components regression (PCR) but partial least squares (PLS) regression performs better than PCR since PCR components are obtained irrespective of the response. Clusterwise PLS regression may be used when heterogeneity in the data is present (Preda *et al.*, 2005). This corresponds to a mixture of several regression models. Clusters and local models are found by an extension of k -means clustering. Previous methods are easily generalized to binary classification, since Fisher's linear discriminant function is equivalent to a multiple regression (Preda *et al.*, 2007). Logistic regression has also been generalized to functional predictors (Escabias *et al.*, 2007). In many real time applications like industrial process, it is of the highest interest to make anticipated predictions. A method based on a bootstrap test for comparing areas under ROC curves helps to determine an optimal time $t^* < T$ giving a prediction based on $[0; t^*]$ almost as good as the prediction using all the data. (Costanzo *et al.*, 2006). Instead of using the same anticipated decision time t^* for all data, we may adapt t^* to each new trajectory given its incoming measurements. The procedure is similar in its spirit to a on-line sequential test (Costanzo *et al.*, 2008). Applications on real and simulated data will be presented.

References

- Costanzo D., Preda C., Saporta G. (2006). Anticipated prediction in discriminant analysis on functional data for binary response. In *COMPSTAT2006*, 821-828, Physica-Verlag
- Escabias M., Aguilera A.M., Valderrama M.J. (2007) Functional PLS logit regression model. *Computational Statistics & Data Analysis*, 51, 10, 4891-4902
- Fisher R.A. (1924) The Influence of Rainfall on the Yield of Wheat at Rothamsted. *Philosophical Transactions of the Royal Society*, B: 213: 89-142
- Preda C. & Saporta G. (2005) Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 49(1): 99-108, 2005.
- Preda C., Saporta G., Lévêder C., (2007) PLS classification of functional data, *Computational Statistics*, 22(2), 223-235
- Ramsay J.O. & Silverman B. (1997), *Functional data analysis*, Springer

DISCOUNTED PAYMENTS THEOREMS FOR LARGE DEVIATIONS

Leonas SAULIS, Dovile DELTUVIENE
 Vilnius Gediminas Technical University, Lithuania
 E-mail: doviled@mail.lt

Let we have a compound Poisson process and a specific random variable. The latter random variable represents the sum of the discounted payments. We obtain large - deviation theorems, exponential inequalities, and a nonuniform estimate in the Berry - Esseen inequality for the properly centred and normed random variable.

STATISTICAL INFERENCE OF QUADRATIC RĒNYI ENTROPY

Oleg SELEZNJEV, Nikolaj LEONENKO
 Umea University, Sweden
 E-mail: oleg.seleznjev@math.umu.se

Entropy and its various generalizations are widely used in mathematical statistics, communication theory, physical and computer sciences for characterizing the amount of information in a probability distribution. The descriptive properties of a record distribution's entropy in a random database are considered in Seleznev and Thalheim (2008). A certain class of estimators of Renyi entropy are studied in Leonenko *et al.* (2008). We propose new estimators of the quadratic Renyi entropy and some related characteristics of

probability distributions based on the number of coincident (or epsilon-close) records in the corresponding i.i.d. sample. We show the asymptotic properties of these estimators (e.g., consistency and asymptotic normality). These estimators can be used in various problems in computer science and mathematical statistics (e.g., average case analysis for random databases; the height of digital trees in information theory; approximate pattern matching in bioinformatics; small ball probability problems).

[1] Seleznev, O. and B. Thalheim (2008). Random databases with approximate record matching. *Method. Computing Appl. Prob.* (online first)

[2] Leonenko, N., Pronzato, L. and V. Savani (2008). A class of Renyi information estimators for multidimensional densities. *Ann. Statistics*, 36, 2153-2182.

MONTE CARLO METHODS FOR PARAMETER ESTIMATION IN NONLINEAR DIFFUSIONS

Yuan SHEN, Dan CORNFORD, Cedric ARCHAMBEAU, Manfred OPPER
Aston University, Birmingham, United Kingdom
E-mail: sheny2@aston.ac.uk

In this paper, we present two Monte Carlo algorithms for parameter estimation in diffusion processes. Both methods compute the marginal likelihood profile. The first method is to make use of so-called Wang-Landau algorithm which has been developed for computing the normalising constant of complex models. Given a lattice in parameter space, the algorithm carries out path sampling for the same process but with different parameter values from the lattice in a dynamical and adaptive way. The second method is based on a variational approximation algorithm which has been recently developed for approximate inference in diffusion processes. With the gradients provided by that algorithm, one can sample the continuous parameter space efficiently. This results in an approximate marginal likelihood profile which is a lower bound of the exact one. Both methods are validated on stochastic double-well system, SINE drift diffusion, and stochastic Lorenz system. Their performance is analysed and discussed when compared to a full Bayesian Markov Chain Monte Carlo method proposed by Golightly and Wilkinson.

ENHANCED SUPERVISION OF AUTOMATIC TELLER MACHINES VIA AUTOASSOCIATIVE NEURAL NETWORKS

Rimvydas SIMUTIS, Darius DILIJONAS, Lidija BASTINA
Vilnius University Kaunas Faculty of Humanities, Lithuania
E-mail: Rimvydas.Simutis@ktu.lt

This paper proposes a new procedure to supervise automatic teller machines' (ATM) network and to detect unexpected behaviour of the ATMs. Early detection of the unexpected behaviour of the automatic teller machines is crucial for efficient functioning of ATM networks. The proposed procedure is based on autoassociative neural networks which form nonlinear correlation models of the input data. Inputs are moving window data of ATM's daily money withdrawals. Autoassociative neural networks have been created for ATM clusters, with similar behaviour. Modelling errors of the networks are used as indicators for detection of unexpected behaviour of ATMs. The proposed procedure has been tested using simulations studies and real experimental data. The simulation results and experimental tests have been shown the efficiency of the proposed procedure. Currently the proposed procedure is being implemented in professional software for supervision and control of ATM networks.

DYNAMIC MODELING OF LIFE TABLE DATA OF EU COUNTRIES

Charilaos Skiadas¹, George Matalliotakis², and Christos H. Skiadas²

¹Hanover College, Indiana, USA

E-mail: skiadas@hanover.edu

²Technical University of Crete, Chania, Crete, Greece

E-mail: gmatal@mycosmos.gr, skiadas@ermes.tuc.gr

In this article a methodology of stochastic processes is performed which is analyzing the life data of the populations of European Union's countries. A related report has been done in which the "health state function" is estimated. Old and new models are applied and the parameters are estimated by means of a non-linear regression analysis algorithm. The study shows that the tested models provide reliable results for life-table analysis. Important conclusions are extracted by testing the different characteristic parameters proposed expressing the mortality of the two genders male and female. The findings concern the increase of the mean life span and the change of "children's mortality" in various countries of European Union.

FIRST EXIT TIME MODELING OF LIFE TABLE DATA

Christos H. Skiadas¹ and Charilaos Skiadas²

¹Technical University of Crete, Chania, Crete, Greece

²Hanover College, Indiana, USA

E-mail: skiadas@ermes.tuc.gr, skiadas@hanover.edu

The growth process of the expected life time of a population is presented and modelled by using several scientific tools. An historical review starts from the first deterministic models mainly due to De Moivre (1724), Gompertz (1825), Weibull and continues with the stochastic models due to Janssen and Skiadas (1995) and Weitz and Fraser (2001). Janssen and Skiadas (1995) proposed a general first exit time model to express the life table data process and several years later new model developments appear in Skiadas (2006) and Skiadas and Skiadas (2007). An application based on a simple model of first exit time is due to Weitz and Fraser. The model they suggest is a special case of that proposed few years earlier by Janssen and Skiadas (1995). We present and analyse dynamic models expressing the human life table data by using the first-passage-time theory for a stochastic process. We present an analytic derivation of some models by applying inverse Laplace transforms and then the models are applied to mortality data. Furthermore the implications of stochastic modeling and the results derived to actuarial science are discussed.

SOLVENCY II - THE NEW PERIOD IN EUROPEAN SYSTEM OF INSURANCE

Jarmila SLECHTOVA

Silesian University in Opava, Czech Republik

E-mail: slechtova@opf.slu.cz

The insurance company Solvency I means an ability of these legal entities to fulfil the liabilities received on the client's contracting relation basement, which can be a legal entity and also a physical entity. These liabilities are settlements of competent insurable titles from insurable events. The pertinent insolvency of an insurer takes place when his assets are insufficiently high or are not insufficiently liquid. It means that it is not possible to covered premium affairs by these assets totally.

Every European Countries are preparing to receive a new system - Solvency II. It is a new regulatory regime designed to provide a principles-based supervisory framework for European insurers and re-insurers. It will also bring an increased level of transparency and harmonisation to the insurance sector like the one part of financial sector.

This new system can bring many problems for the small insurance companies for instance their basic capital and administration costs.

EXACT ERROR RATES OF THE SUPERVISED CLASSIFICATION BASED ON MARKOV RANDOM FIELDS

Lijana STABINGIENE, Kestutis DUCINSKAS
Klaipeda University, Lithuania
E-mail: kestutis.ducinskas@ku.lt

In spatial classification it is usually assumed that feature observations given class labels are independently distributed. We have restrictive assumption considering stationary Gaussian field model for feature observations. The class label vector is assumed to follow Markov Random Field model. Formulas for exact error rates are derived in case of isotropic exponential spatial correlation. The dependence of error rates on the values of statistical parameters of considered models is investigated numerically.

EVOLUTION OF A CREDIT PORTFOLIO USING THE CREDITMETRICS METHOD AND SIMULATION TECHNIQUES AT THE BANK LEVEL

Stelian STANCU, Gabriela SAVA, Iuliana GEGEA
Academy of Economic Studies, Romania
E-mail: stelian_stancu@yahoo.com

Credit risk is generally defined as the risk of financial loss triggered by the counterpart's not having paid the amounts due. Its effect can be measured by the costs incurred by replacing the financial flows that may have occurred should the counterpart have met their obligations. As bankers around the world entities have been granting more and more credits exposed to risk, risk credit has become the main challenge in the management field. Credit risk management is a special kind of risk as it has a wider scope than a simple loan operation. The former expands and encompasses such background aspects related to the three stages of the crediting process, namely: the decision making as to crediting, the credit life process, and the credit portfolio management. This paper focuses on the Bank credit portfolio management. We use the CreditMetrics quantitative model as an instrument to foresee the exposure of the entire portfolio to credit risk. This model assesses the bank portfolio making estimates as to risk exposure due to changes undergone by the value of the portfolio as a result of changes likely to occur as to the quality of the credits granted to debtors. The changes the quality of the credits may undergo imply both non-payment events and increases or decreases of the quality of the credits granted. Thus, the fundamental source of the risk is the date on which the quality of a debtor's credit modified against the risk horizon as well as the likeliness that the debtor may not pay, may not prove to be good willed or a successful attempt at fraud.

DENSITY ESTIMATOR OF COARSE DATA

Eugenia STOIMENOVA
Institute of Mathematics and Informatics, Bulgaria
E-mail: jeni@math.bas.bg

Coarse data refer to data with low quality. We run into this type of data in situations such as missing data, censor data (say, in survival analysis), and grouped data. It is the difficulty in observing accurate data that leads to coarse data. This might happen in performing random experiments as well as observing random phenomena in nature. Suppose our random variable X of interest takes values in U . Suppose X cannot be observed directly but its outcomes can be located in a finite partition τ_1, τ_2, \dots , which is independent of X . A such partition is referred to as a coarsening of X . Define $[R = \inf \tau_j : \tau_j \geq X, L = \sup \tau_j : \tau_j \geq X]$. For each X_i the corresponding interval Y_i may be defined on a different partition of X . The conditional density is itself unknown. The general aim of exploratory data analysis is to find interesting features in data, and to bring them to human perception. Probability density estimation (usually identify with its plot) serves for these purposes. This paper is concerned with the nonparametric estimation of a density function when the data are incomplete due to coarseness. The histogram density estimator is modified to allow description of such coarse data. An interactive R application is developed to explore different estimates.

EXCEEDANCE-TYPE TEST AGAINST LEHMANN ALTERNATIVES

Eugenia STOIMENOVA
Institute of Mathematics and Informatics, Bulgaria
E-mail: jeni@math.bas.bg

Rank tests based on the number of observations from one sample which exceed some random level allow simple comparison of two distributions. Let X_1, \dots, X_m and Y_1, \dots, Y_m be two independent samples from two distributions with continuous distribution functions F and G respectively. The Standard problem is that of testing the hypothesis H_0 that F and G are identical against the one-sided alternative H_A that X 's are stochastically larger than Y 's: $G(x) \geq F(x)$, with strict inequality for some x .

Define A and B' to be the number of observations among X_1, \dots, X_m larger than $\max_{1 \leq j \leq n} Y_j$, or smaller than $\min_{1 \leq j \leq n} Y_j$, and A' , and B to be the number of observations among Y_1, \dots, Y_m larger than $\max_{1 \leq j \leq n} X_j$, or smaller than $\min_{1 \leq j \leq n} X_j$. The simplest test for H_0 against H_A is due to Rosenbaum (1954) and is based simply on the statistic A , i.e. the number of observations from X -sample that exceed all observations from the Y -sample. Clearly, large values of A lead to the rejection of H_0 . This test suffers from a masking effect of the number of observations from Y -sample that precede all observations from the X -sample, the statistic B . Large values of B should also lead to the rejection of H_0 . In order to avoid this masking effect, we propose the following test statistic for this hypothesis testing problem $M = \max \{n-A', m-B'\}$. The null hypothesis is rejected in favor of H_A for small values of M .

The M statistic is proposed by Stoimenova (2000) using a distance based approach that produces many popular rank tests and gives rise some new. This two-sample test is a particular member of a family of rank test induced by Chebyshev's metric on permutations. The exact distribution of the M -statistic under the null hypothesis is distribution free and is derived in the same paper. Difficulty arises in determining the power of non-parametric tests due to the generality of the alternative H_A . The present paper deals with the distribution of the M -statistic under Lehmann alternative H_{LE} : $G(x) = 1 - (1 - F(x))^{1/\eta}$, for some $\eta > 1$. This subclass of H_A was introduced by Lehmann (1953), and clearly, it can be considered as a rough approximation for a shift of G to the right with respect to F . We obtain that the distribution of M under H_{LE} is distribution free and is given by an explicit expression. The distribution under the alternative H_{LE} can be used for calculating the power of the test, i.e. the probability of rejecting the null hypothesis when the alternative hypothesis is indeed true. We compare the power of the M -test with the power of some other rank tests for the two-sample testing problem given by Balakrishnan and Frattina (2000), and Balakrishnan et al. (2008).

MULTI-POPULATION GENETIC ALGORITHM FOR LOCATING MULTI-OPTIMA IN COMPLEX LANDSCAPE

K.Y. SZETO, Yunbo GUO
Chinese University of Hong Kong
E-mail: phszeto@ust.hk

Evolutionary computation based on ideas of spatial-temporal resource allocation is developed under the framework of a multi-agent system consisting of sub-populations of agents. These agents evolve simultaneously as group of chromosomes in a genetic algorithm. The multi-agent system becomes a parallel genetic algorithm, while the spatial allocation of resource is achieved by the partitioning of the search space into many subspaces. Each sub-population contains a set of agents undergoing evolution with mutation and crossover using the recently developed mutation matrix formalism. A migration operator is used to control the exchange of chromosomes between different sub-populations. This spatial allocation of computational resource takes the advantage of exhaustive search which avoids duplicate effort, and combine it with the parallel nature of the search for solution in disjoint subspaces by genetic algorithm. The division of the solution space is performed intelligently using loci statistics of the chromosomes in past generations. The time when this division takes place is determined by monitoring the performance of the evolutionary computation using mean and variance. The temporal allocation of resource is formulated by an adaptive evolutionary algorithm that combines two given genetic algorithms using an adaptive resource allocation of

CPU time, based on ideas of Markowitz investment frontier and the Sharpe ratio. We illustrate these general ideas using the example of finding the local optima of a complex function. For a benchmark function with known number of optima, we can evaluate the performance of our multi-population genetic algorithm by two criteria: the percentage of the optima found, and the error to the true optima. A good landscape mapping algorithm will optimize both criteria with minimum CPU time. Numerical experiments were performed on a set of benchmark functions. Our results on benchmark functions show that MPGA provides a good tool for landscape mapping. Application of our method to map out complex landscape of functions that cannot be written down explicitly, such as protein folding, is suggested.

LOGISTIC SYSTEM BASED ON AGENTS CLUSTERING USING FUZZY AND NEURAL METHODS

Ioan Constantin TARCA, Tiberiu VESSELENYI, Florin BLAGA,
Lehel Szabolcs CSOKMAIL, Radu Catalin TARCA
University of Oradea, Romania
E-mail: rctarca@yahoo.com

Transport optimization issues are in the focus of many studies, which concerns costs minimization, maximization of transported cargos, minimization of delivery time. Although there are a lot of internet based systems which facilitates transport service searching very few of them are integrating transport optimization algorithms. Our study presents some solutions regarding transport optimization based on clustering of transport service providers and clients using fuzzy and neural methods. These algorithms are meant to be implemented in an internet based system which can be used both by clients and transporters in order to maximize their profits.

FORECASTING CPI USING A NEURAL NETWORK WITH FUZZY INFERENCE SYSTEM

Camelia Ioana UCENIC, George ATSALAKIS
Technical University Cluj-Napoca, Romania
E-mail: cameliaucenic@yahoo.com

Consumer price index (CPI) is an index of the cost of all goods and services of a typical consumer. CPI is a key measure of inflation that relates the rise in prices over a period of time. It is a fixed quantity price index and is used as an economic indicator, more specific as an inflationary indicator. As the most widely used measure of inflation, CPI is an indicator of the effectiveness of government policies. The business executives, labour leaders and private citizens use the index as a guide in making economic decisions. It is also called cost-of-living index. This work presents an attempt to forecast CPI using an adaptive model, a neural network with fuzzy inference system. The adaptive models transform the internal behavior structure to include changes in the outside world. The data used for the model is related to Greece and is from OECD Factbook 2008. The value of the year 2000 is consider for comparison and is equaled with 100. The results are compared with traditional forecasting models.

FUZZY MODEL APPLIED TO FORECAST THE PROMOTIONAL EXPENSES

Camelia Ioana UCENIC, Laura BACALI, Carmen TRIF
Technical University Cluj-Napoca, Romania
E-mail: cameliaucenic@yahoo.com

The firms are trying to enlarge their sales through promotion. This is obvious from the increase in the share of promotion expenses of private sector companies. Nowadays experience argues about an amount of potential cost advantages and shortcomings. The expenditure in promotional marketing sub-mix is among them. The promotional efforts of a brand with small consumption rates can convert the brand into a challenger for the well-liked brands. A drop of promotional expenditure is reasonable when products are no longer promotionally responsive. If this is not the situation, a decrease of promotional spend from existing

rank is possible to direct to losses in market share and sales that significantly are more important than the savings in promotional expenditure.

This study uses a fuzzy model in order to forecast the promotional expenditure using its previous values. This research provides some information to guide future managerial decisions among the available opportunities.

MULTI-STAGE SEMI-MARKOV MODELS FOR RECOVERY PROCESS

Fidelis UGWUOWO, Francis UDOUMOH
University of Nigeria, Nsukka, Udoumoh
E-mail: fiugwuowo@yahoo.com

Some models for the recovery process of cataract patients after surgery are presented. The use of semi-Markov modeling approach is adapted in a multi-stage process involving acuity levels of patients. The Parametric and Non-parametric methods of estimation were used for Maximum likelihood estimation of the sojourn time distribution. The stationary distribution of the Semi-Markov process (SMP) reveals 6/18 as best corrected acuity with 0.3485 proportions of patients.

ON THE MULTICOLLINEARITY IN NONLINEAR REGRESSION

Mujgan TEZ, Kadri Ulas AKAY, Ali ERKOS
Marmara University, Istanbul, Turkey
E-mail: mtez@marmara.edu.tr

The multicollinearity in nonlinear regression is the subject of this study. If there is no linear relationship between the regressors, they are said to be orthogonal. When the regressors are orthogonal, inferences can be made relatively easily. If the regressors are not orthogonal, the usual least-squares analysis of the regression model is inadequate. There are several methods available in literature for detection of multicollinearity in linear regression but assessing the severity and consequences of multicollinearity in nonlinear regression is more complicated than in linear regression. Complete elimination of multicollinearity is not possible but the degree of multicollinearity can be reduced by adopting some methods. In this study, these methods will be analyzed and examined for given data set.

DEPTH BASED CLASSIFICATION

Ondrej VENCALEK
Charles University in Prague, Czech Republic
E-mail: venco2am@artax.karlin.mff.cuni.cz

Data depth is a modern nonparametric approach to multivariate data analysis. One particular area of applications is a classification (or discrimination). In the talk we show the usual classification rules based on the data depth. Their main advantage is possible relaxing of normality assumption. In particular we study classification under an alternative recently introduced depth functions.

ON APPLICATION OF THE K-NEAREST NEIGHBOURS APPROACH FOR CLUSTER VALIDATION

Zeev VOLKOVICH, Zeev BARZILY, Renata AVROS, Dvora TOLEDANO-KITAI
ORT Braude College of Engineering, Israel
E-mail: zeev@actcom.co.il

K-Nearest Neighbors (K-NN) is a widely used technique for classifying and clustering data. However, probabilistic aspects of this methodology are typically not involved in applications. In the current paper, we address the cluster stability problem resting upon probabilistic characteristics of the (K-NN) approach. For

several possible numbers of clusters we estimate the stability of partitions obtained from clustering pairs of samples such that partitions are presumed to be consistent if their clusters are stable. Clusters validity is quantified through the amount of K-NN belonging to the point's sample. The null-hypothesis, of the well-mixed samples within the clusters, suggests Binomial Distribution of this quantity with K trials and the success probability $p_0=0.5$. A cluster is described by the mean of the p-values calculated over all cluster points within testing the hypothesis $p_0=0.5$ against the alternative $p_0>0.5$. Consequently, the partition quality is measured via the worst partition cluster corresponding to the maximal mean value. The true number of clusters is attained by the empirical distribution of this mean value having the shortest right tail. The proposed methodology offers to produce the mean-value distributions sequentially and to assess their right-asymmetry. Numerical experiments, presented in the paper, exhibit the capability of the methodology to expose the true number of clusters.

THE EXTENDED HYPERBOLIC SMOOTHING CLUSTERING METHOD

Adilson Elias XAVIER, Vinicius Layter XAVIER
Federal University of Rio de Janeiro, BRAZIL
E-mail: adilson@cos.ufrj.br

The minimum sum-of-squares clustering problem is considered. The mathematical modeling of this problem leads to a $\min\text{-sum}\text{-min}$ formulation which, in addition to its intrinsic bi-level nature, has the significant characteristic of being strongly nondifferentiable. To overcome these difficulties, the proposed resolution method, called Hyperbolic Smoothing, adopts a smoothing strategy using a special C^∞ differentiable class function. The final solution is obtained by solving a sequence of low dimension differentiable unconstrained optimization subproblems which gradually approach the original problem. This paper presents an extended method based upon the partition of the set of observations in two non overlapping parts. This last approach engenders a drastic simplification of the computational tasks.

THE HYPERBOLIC SMOOTHING APPROACH FOR SOLVING A SUPPORT VECTOR MACHINE PROBLEM: NEW IMPROVEMENTS

Adilson Elias XAVIER, Victor Stroele de Andrade MENEZES,
Patricia Curvelo RODRIGUES
Federal University of Rio de Janeiro, Brazil
E-mail: adilson@cos.ufrj.br

This paper is intended to show a new approach, HSSVM, for solving a linear Support Vector Machine problem formulation that leads to the minimization of a completely differentiable problem in a space of dimensionality equal to that of the input space. The resolution method proposed adopts a smoothing strategy using a special completely differentiable class function. It is discussed new procedures in order to offer improvements on the performance on the HSSVM.

SUPERCritical CATALYTIC BRANCHING RANDOM WALKS

Elena YAROVAYA
Moscow State University, Russia
E-mail: yarovaya@mech.math.msu.su

We consider two models of continuous-time random walk on multidimensional lattices with birth and death of particles at a single site (i.e., the source of branching). The main difference between the models lies in the behavior of the process at the source. Model I assumes that the random walk is symmetrical. In Model II it is entered an additional parameter enabling artificial intensification of prevalence of branching or walk at the source and, as the result, violating symmetry of the random walk. In Model II the differential equations for the generating functions of the numbers of particles both at an arbitrary point of the lattice and

on the entire lattice were established. Sufficient conditions for exponential growth of the process are established. General methods to study the models in the supercritical case are proposed.

IRREGULAR SAMPLING IN THE ESTIMATION OF TRANSFER FUNCTIONS

Vladimir ZAIATS
Universitat de Vic, Spain
E-mail: vladimir.zaiats@uvic.es

Irregular sampling is a useful tool in the estimation of correlation functions and spectral densities in continuous-time stochastic processes. For example, periodic sampling of a process of this type is well-known to introduce the so-called aliasing of the spectrum. The first notion of alias-free sampling was developed by Shapiro and Silverman in 1960. Masry introduced an alternative definition in 1978. Then he obtained alias-free spectral estimates based on irregular sampling, whose rates of quadratic-mean convergence are identical to those of the estimates based on continuous-time observations.

We extend the ideas of irregular sampling as applied to the estimation of the transfer function of a general linear time-invariant system. Our estimate is based on the sample cross-correlogram between the input and output. We study asymptotic properties of the estimate and compare them with those obtained under periodic sampling.

CANONICAL MEASURES OF A COLORED DISORDERED LATTICE GAS

Halim ZEGHDOUDI, Hacène BOUTABIA
LANOS laboratory, Badji-Mokhtar University, BP12, Annaba 23000-Algeria
E-mail: hzeghdoudi@yahoo.fr

In this work we deal with grand canonical and canonical measures related to a model called colored disordered lattice gas. We consider the approach used in the work of A. Dermoune and P. Heinrich [1]. We prove a new estimate for the canonical measures and we made a generalization in the multi-coloured case for hydrodynamic limit.

SPECTRAL GAP FOR MULTICOLOR DISORDERED LATTICE GAS OF EXCLUSION PROCESSES

Halim ZEGHDOUDI, Ali bey TOUATI, Hacène BOUTABIA
LANOS laboratory, Badji-Mokhtar University-BP12, Annaba 23000- Algeria.
E-mail: hzeghdoudi@yahoo.fr

We consider a system of multicolored disordered lattice gas in a volume Λ of \mathbf{Z}^d : Using a result of Caputo [3], who prove spectral gap inequalities in product space with conservation laws of generalized exclusion and zero range processes. The aim of the present paper is the estimate the spectral gap for multicolored disordered lattice gas of exclusion processes which plays an important role in the study of hydrodynamic limit.

A HYBRID METHOD FOR ESTIMATING PARAMETERS OF WIENER PROCESS AND NOISE FROM NOISY OBSERVATIONS.

Antanas ZILINSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: antanasz@ktl.mii.lt

It is supposed that observations of Wiener process values in the presence of noise are available. The variance parameters of the process and of noise should be estimated. Numerical implementation of maximum likelihood method is difficult because of necessity to invert large ill defined matrices. In the

present paper a hybrid method is proposed combining ideas of methods of maximum likelihood and of least squares. Results of simulation are included.

GLOBAL OPTIMIZATION ORIENTED STOCHASTIC INTERVAL ARITHMETIC

Antanas ZILINSKAS, Julius ZILINSKAS
Institute of Mathematics and Informatics, Lithuania
E-mail: zilinskasjulius@gmail.com

Recently two randomized versions of interval arithmetic under the title ‘random interval arithmetic’ were proposed and applied to global optimization. In the present paper some drawbacks of random interval arithmetic are discussed, and improvements are proposed to exclude the mentioned drawbacks. The newly proposed (stochastic interval) arithmetic is experimentally compared with the previously published version of random interval arithmetic.

Authors List

ABBASSI Noufel 56
ADEYEMI Shola 4
AFANASEVA Larisa 4
AGUILERA Ana M. 16
AKAY Kadri Ulas 4, 68
ALZBUTAS Robertas 5, 42
ANDRONOV Alexander 7
ANTAL Raluca Meda 7, 48
ANTONOV Valery 7
ARATO Miklos 8
ARCHAMBEAU Cedric 63
ARIAS-NICOLES J.P. 8
ASIKGIL Baris 9
ATSALAKIS George 67
AVROS Renata 68
BACALI Laura 67
BAGIROV Adil 9
BAGKAVOS Dimitrios 10
BAKSYS Donatas 10
BALAKRISHNAN Narayanaswamy 10, 35
BARDEL Noemie 11
BARDET Jean-Marc 11
BARTKUTE-NORKUNIENE Vaida 11, 12
BARTON Maria 12
BARZILY Zeev 68
BASTINA Lidija 63
BAZIUKAITE Dalia 13
BECUE-BERTAUT Monica 13
BELOVAS Igoris 36
BERTRAND Pierre R. 11
BIFFI Elena 14
BLAGA Florin 67
BLOCK Henry W. 14
BOCCHI Leonardo 30
BOGDANOVICH Veniamin 14
BORISENKO Ingrida 24
BOUGEARD Stephanie 15
BOUTABIA Hasane 70
BOUVEYRON Charles 15
BOUZAGHTI Yamina 40
BOUZAS Paula R. 16
BOYACIOGLU Melek Acar 16
BRAUMANN Carlos A. 28
BRUNET Camille 15
BULINSKAYA Ekaterina 4
CADORET Marine 16
CANO Javier 17
CARSTEA Horia 50
CHAN P.S. 17, 35
CHAUSSALET Thierry 4
CHAVENT Marie 18
CHENARU Oana 18
CIUMARA Roxana 58, 59
COLMENARES Berenice 13
CONDE-SANCHEZ Antonio 53, 60
CORNFORD Dan 63
CRISTEA Enona 18
CRISTEA Lidia 18
CSOKMAI Lehel Szabolcs 67
CUVAK Ana 19
CUVELIER Etienne 19
D'AMICO Guglielmo 7
DABALA Wieslawa 20
D'AMATO Valeria 5
DANILENKO Svetlana 20
DANILEVICH Janush 7
DASH Gordon 20
DEDU Silvia 59
DELSOL Laurent 21
DELTUVIENE Dovile 62
DEMIR Eren 4
DESBOUVRIES Francois 11
DEVOLDER Pierre 21
DILIJONAS Darius 63
DINU Liviu 22
DOLEV Shlomi 23
DONCHEV Doncho 23
DRAGUT Andreea 23
DRAGUT Maria 23
DREIZIENE Lina 24
DUCINSKAS Kestutis 24, 65
DURIO Alessandra 25
DZEMYDA Gintautas 25, 38
ELEFThERIOU Mavroudis 25
ELLIS Tim 39
ENACHESCU Cornelia 26
ENACHESCU Denis 26
ERAR Aydin 9
ERDOGAN Birsen Eygi 26
ERKOS Ali 68
ESCABIAS Manuel 52
FARIA Susana 26
FARMAKIS Nicolas 25
FERNANDEZ-AGUIRRE Kar mele 27
FERNANDEZ-ALCALA Rosa M. 27
FERRATY Frederic 21
FHIMA Medhi 11
FILIPE Patricia A. 28
FILUS Jerzy 28
FILUS Lidia 28
FLEURY Gilles 43
FLOREA Gheorghe 18
FOGLIATTO Flavio 51
FORMAN John 20
FRANCO Manuel 29
FREEMAN James 29

FRENKEL Sergei 23
 FULGA Cristina 29
 FULLERTON Ken 12
 GAMIZ M.Luz 43
 GARG Lalit 12, 29
 GARRETA Henri 33
 GEGEA Iuliana 65
 GIACOMO Paola Di 30
 GIEBEL Stefan 30
 GIUSEPPE Di Biase 31
 GOLET Ioan 31
 GOMES Dulce 32
 GOMES M. Ivette 31
 GREENGLAZ Leonid 40
 GRUSHO Alexander 32
 GUENOCHE Alain 33
 GUERMEUR Yann 33
 GUGLIELMO D'Amico 31
 GUILLEN Montserrat 7
 GUILLET Fabrice 46
 GULATI Sneh 33
 GULER Nesrin 33
 GUO Yunbo 66
 HABART Marine 34
 HABERMAN Steven 5
 HE Qi-Ming 34
 HEDREA Ciprian 50
 HILLION Alain 34
 HKUST 44
 HRUSCHKA Harald 34
 IATAN Iuliana 35
 INSUA David Rios 17
 ISAIA Ennio Davide 25
 IVANIKOVAS Sergejus 25
 JACQUES Jannssen 31
 JAKIMAUSKAS Gintautas 35
 JANSSEN Jacques 14
 JUSCENKO Natalija 13
 KA C.Y. 35
 KAASIK Ants 36
 KABASINSKAS Audrius 36
 KADANKOVA Tetyana 36
 KAJIJI Nina 20
 KAKLAUSKAS Liudvikas 37
 KAKUBAVA Revaz 37
 KALAMARAS Dimitios 41
 KALAMATIANOU Aglaia 10, 41
 KALINAUSKAS Zilvinas 19
 KARAKOS Alexandros 38
 KARAKOS Pericles 38
 KARBAUSKAITE Rasa 38
 KATYSHEV Pavel 38
 KAVALIERIS Laimonis 39
 KAYA Ahmet 39
 KEMPA Wojciech 39
 KHALDI Khaled 40
 KHARIN Yuriy 40
 KOLLO Tonu 55
 KONG Honk 44
 KONSTANTINIDES Dimitrios 40
 KOPYTOV Eugene 40
 KOUGIOUMOUTZAKI Foteini 41
 KOVALENKO Anatoly 7
 KUBILIUS Kestutis 41
 KUDZYS Antanas 41
 KUENTZ Vanessa 18
 KUPCIUNIENE Kristina 42
 KURASOVA Olga 42, 43
 KUTTYKRISHNAN A.P. 43
 LAGARTO Sandra 32
 LAHALLE Elisabeth 43
 LANDALUCE-CALVO M. Isabel 27
 LARA-PORRAS Ana 43
 LE Sebastien 13, 16
 LEDOIT Olivier 44
 LEEN Seral 16
 LEMESHKO Boris 44
 LEMESHKO Stanislav 44
 LENCA Philippe 34
 LEONENKO Nikolaj 62
 LIANG Tianzhu 44
 LIN C.T. 17
 LIQUET Benoit 45
 LOMBARDI Luigi 45
 LORENZO Emilia Di 5
 LUKOSEVICIENE Ona 41
 MABROUK Mohamed Ben 56
 MAK M.Y. 17
 MALINOVSKII Vsevolod 45
 MALUGIN Vladimir 46
 MANCA Raimondo 7, 14, 31
 MANITA Anatoly 46
 MARINICA Claudia 46
 MARKUS Laszlo 46
 MARSHALL Adele 47
 MARTINEZ-RODRIGUEZ Ana Maria 53, 60
 MARY-HUARD Tristan 47
 MATALLIOTAKIS George 64
 MAUCERI Christian 48
 MAUMY-BERTRAND Myriam 48
 MCCLEAN Sally 12, 29
 MEDVEDEV Viktor 25
 MEENAN Brian 29
 MELICHOV Dmitrij 41
 MENDONCA Sandra 31
 MENEZES Victor Stroele de Andrade 69

MEYER Nicolas 48
 MIHALCA Gabriela Maria 7, 48
 MILLARD Peter 29
 MINKOVA Leda 49
 MODROS-HERREN Juan Ignacio 27
 MOLINA Manuel 49
 MOLYTE Alma 43
 MONFRINI Emmanuel 33
 MORENO Alberto 49
 MOTA Manuel 49
 MURAVJOVS Aivars 40
 NAVARRO-MORENO Jesus 27
 NEDELKO Svetlana 49
 NEDELKO Victor 50
 NEGREA Romeo 50
 NEUBAUER Jiri 50
 NG H.K. 17, 35
 NIANG Ndeye 51
 NIGLIO Marcella 51
 NOIRHOMME-FRAITURE Monique 19
 NORVAISA Rimas 52
 NUNES Carla 32
 OBEDA Xavier 54
 OCANA-LARA Francisco A. 52
 OCANA-PEINADO Francisco M. 52
 OCHEANA Luiza 18
 OLIVA Marc 52
 OLMO-JIMENEZ Maria Jose 53, 60
 OPPER Manfred 63
 OTTO Wojciech 53
 OZDEMIR Halim 33
 PAGES Jerome 16
 PAPANASTASSIOU Demetrios 54
 PARNA Kalev 36
 PASTORE Massimiliano 45
 PEREIRA Paulo 52, 54
 PEREZ-OCON Rafael 55
 PESTANA Dinis 31
 PETTERE Gaida 55
 PIECZYNSKI Wojciech 55, 56
 PISCOPO Gabriella 56
 POLAT Kemal 16
 POPESCU Marius 22
 PREDĂ Ciprian 50
 PREDĂ Cristian 58
 PREDĂ Vasile 58, 59
 PROKAJ Vilmos 59
 QANNARI El Mostafa 15
 RAMOS Alfonso 49
 REYNES Christelle 59
 ROBIN Stephane 47
 RODRIGUES Patricia Curvelo 69
 RODRIGUEZ-AVI Jose 53, 60
 ROMAN-MONTOYA Yolanda 43
 ROQUETE Carlos J. 28
 RUIZ-FUENTES Nuria 16
 RUIZ-MOLINA Juan C. 27
 RUKSENAITE Jurga 60
 RUSSOLILLO Maria 5
 SAEZ-CASTILLO Antonio Jose 60
 SABATIER Robert 59
 SAKALAUŠKAS Leonidas 10, 12, 36, 37, 61
 SALTYTE Laura 61
 SANTALOVA Diana 7
 SAPORTA Gilbert 51,61
 SARACCO Jerome 18, 45
 SAULIS Leonas 62
 SAVA Gabriela 65
 SAVITS Thomas H. 14
 SCHENK Jens-Peter 30
 SCHILTZ Jang 30
 SEGOVIA M. Carmen 55
 SELEZNJEV Oleg 62
 SHAPIRO Samuel 33
 SHEN Yuan 63
 SIBILLO Marilena 5
 SIMUTIS Rimvydas 63
 SKIADAS Charilaos 64
 SKIADAS Christos H. 64
 SLECHTOVA Jarmila 64
 SOROMENHO Gilda 26
 SOUZA Michael 49
 STABINGIENE Lijana 65
 STANCU Stelian 65
 STOIMENOVA Eugenia 65, 66
 SZETO Kwok Yip 44, 66
 TARCA Ioan Constantin 67
 TARCA Radu Catalin 67
 TEZ Mujgan 33, 68
 TICHIT Laurent 33
 TIMONINA Elena 32
 TOLEDANO-KITAI Dvora 68
 TOUATI Ali bey 70
 TRIF Carmen 67
 UCENIC Camelia Ioana 67
 UDOUMOH Francis 68
 UGON Julien 9
 UGWUOWO Fidelis 68
 URGAN Nurkut Nuray 39
 VAICIULYTE Ingrida 61
 VALDERRAMA Mariano J. 16, 52
 VENCALĖK Ondrej 68
 VERAVERBEKE Noel 36
 VESELY Vitezslav 50

VESSELENYI Tiberiu 67
VIEU Philippe 21
VITALE Cosimo Damiano 51
VIVO Juana-Maria 29
VOLCHEZ-LOPEZ Silverio 53
VOLKOVICH Zeev 32, 68
VOSTRETSOV Aleksey 14
WANG Jie 14
WEBB Dean 9
WOLF Michael 44
XAVIER Adilson Elias 49, 69
XAVIER Vinicius Layter 69

XIE Jingui 34
YAROVAYA Elena 69
YIP Kwok 44
ZAGAJNOV Artem 7
ZAIATS Vladimir 70
ZEGHDOUDI Halim 70
ZENGA Mariangela 47
ZGHEIB Rawad 43
ZHAO Xiaobo 34
ZILINSKAS Antanas 70, 71
ZILINSKAS Julius 71