DRAFT

BOOK of ABSTRACTS

ASMDA 2021

and

Demographics 2021 Workshop

http://www.asmda.es/asmda2021.html

Please check your Abstract and respond only if any error occurs.
If your abstract is missing resend it in Word format.
We will use your Abstract Title and Authors names to formulate the Conference Program only if you have send your Registration form to the Secretariat at secretariat@asmda.com

ASMDA 2021

BOOK of ABSTRACTS

PLENARY

Accelerated Life Testing of One-Shot Devices: Data Collection and Analysis

N. Balakrishnan
Distinguished University Professor McMaster University
Hamilton, Canada

In this talk, I will explain the problem of one-shot device testing analyses. After giving some motivating data sets, I will describe different methods of inference and also detail inferential methods and optimal designs under constant-stress (CSALT) and step-stress (SSALT) accelerated tests. Several real-life examples and simulation study results will be presented to illustrate all the models and inferential methods described.

Modeling the development of consonant clusters with singleton consonants in children's speech

Elena Babatsouli
Department of Communicative Disorders, University of Louisiana at Lafayette, USA

Children have more difficulty producing two consonants in sequence, known as consonant clusters, in their speech than one consonant in sequence with a vowel, known as singleton consonants. The accuracies of consonant clusters and of singleton consonants are evaluated separately in practice and it is not known whether the two are correlated. In the present talk such a correlation is investigated based on data from several children's speech samples spanning the accuracy spectrum. Upper and lower bounds of cluster accuracy are obtained analytically in terms of singleton accuracy yielding a maximum deviation of the bounds at 50% singleton accuracy. All possible speech productions fall within these bounds. A positive strong correlation between the two accuracies is found across the whole accuracy spectrum. Even stronger correlation is found if the spectrum accuracy is separated into two regimes by the 50% singleton accuracy. This separation also reveals that although word-final clusters are produced more accurately than word-initial clusters in the lower part of the spectrum, their correlation to singleton consonants is not as strong as word-initial clusters across children with impaired speech.
Keywords: Consonant clusters, Singleton consonants, Children, Normal and impaired speech.

The Conditional Mean and Variance of the Duration of the Game, Given the Barrier First Reached, in the Classical Gambler's Ruin Model

Mark Brown
Department of Statistics, Columbia University, New York, USA

We consider aspects of the classical gambler's ruin problem which goes back to correspondence between Fermat and Pascal between 1654-1656,

with later contributions by Huggens, De Moivre, Nicholas and James Bernoulli, Lagrange and Feller, among many others. The model is that a gambler starts with i dollars and plays a series of games. At each stage the gambler wins 1 unit of currency with probability p, or loses 1 with probability 1 −p. The gambler quits the first time his/her fortune reaches m (m > i), or zero, whichever comes first. Define N to be the duration of the game, and SN the gambler's fortune, m or zero, when the game ends. Converting the problem into an equivalent one involving birth and death processes with a single absorbing state, we derive an expression for the variance of the first passage time to the absorbing state. This holds for a general class of birth and death processes, and in the case of the gambler's ruin model reduces to simple expressions for the condtional mean and variance of N, given SN. We believe that the conditional variance result is new.

Risk Analysis: Classical Statistics compared to Neural Networks in Survival Analysis. Application to Dementia Prediction

Catherine Huber

University of Paris, National Scientific Research Center 8145, MAP5, Biomedical Department, Paris, France

Risk analysis is a topic of increasing importance in multiple fields like environment, technology and biomedicine. In survival data analysis and reliability, one is interested in all risk factors that may accelerate or decelerate the life length of individuals or machines.

In survival, reliability and degradation analysis, classical statistics are based on various probability models[1]: from the simplest to the more sophisticated, parametric models like Weibull and Gamma, semi-parametric models like Cox model and its generalizations and Frailty model, non–parametric models, and latent variable models Threshold regression model, in two versions, parametric and non parametric[2].Now, as immense data bases (big data) are available, several types of methods are needed to deal with the resulting high dimension: on one hand, methods that reduce the dimension while maximizing the infor-mation left in the reduced data, and then applying classical statistical models; on the other hand algorithms that apply directly to big data, i.e. artificial

intelligence (machine learning), at the cost of a difficulty of interpretation in terms of the risk factors. Actually, those algorithms have a probabilistic interpretation. Finally, performances for predicting Dementia are compared for both types of methods[3].

References:

1. Huber, Catherine, Solev, Valentin, and Vonta, Filia. "On Censored and Truncated Data in Survival Analysis and Reliability Models." Stochastic Models in Reliability Engineering (2020): 199.
2. Mei Ling Ting Lee and Whitmore, G.A, “Distribution-free Predictive Inference for Failure Data Using Threshold Regression”, (2021) preprint.
3. Huber, Catherine, Gross, Shulamith and Vonta, Filia, “Risk analysis: Survival data analysis vs. machine learning. Application to Alzheimer prediction. Comptes Rendus Mécanique (2019), vol. 347, no 11, p. 817-830.

Keywords: Dementia, Survival Analysis, Neural Networks.

A chain-binomial epidemic with asymptomatics motivated by COVID-19 modelling

Claude Lefèvre
Université Libre de Bruxelles

Motivated by modelling epidemics like COVID-19, this paper proposes a generalized Reed-Frost process which integrates two types of infectives, those with symptoms and those without. Testing of infectives and vaccination of susceptibles are then incorporated as preventive protective measures. Our interest relates to the distribution of population state at the end of infection and to the basic reproduction number R_0 with the associated extinction condition. The method uses the construction of a family of martingales and a branching approximation for large populations, respectively. A more general branching process for epidemics is also constructed and studied. Finally, some results obtained are illustrated by

numerical examples. This is a joint work with P. Picard (Lyon), M. Simon (Barcelona) and S. Utev (Leicester).

Attaining Process Completion targets based on Markov models and their extensions

Sally McClean
Ulster University, Republic of Ireland

The area of Process Mining has recently emerged to leverage, often voluminous, time-stamped process event data concerning sub-processes or tasks. Such data are often found in diverse topics such as business, industry, telecommunications or healthcare. Process target modelling can be used off-line, or in real-time, to predict whether a predetermined target, or service-level agreement (SLA), is likely to be achieved by a currently running process. This is essentially achieved using a classification model where the extracted features can characterise the structure and nature of the executing process and predict if it is likely to attain the completion target. If not, then mitigating intervention can be performed to minimise damage and disappointment. On the other hand, processes can also use Markov, or related, models to assess compliance with completion targets or, inversely, we can determine appropriate targets for satisfactory performance. In addition, there may be several terminating states from the Markov process, with different distributions of completion times. Results for such set-ups will be illustrated using data from a stroke patient unit, where there are multiple discharge destinations for patients, namely death, private nursing home, or to the patient's own home, where dissimilar discharge destinations may require different targets. Such SLAs are common in healthcare, business, and industrial processes where Markov models, or their extensions, can provide useful tools for process modelling and compliance management.

A Multivariate Concept of Stochastic Precedence and Aggregation-type Paradoxes of Voting Theory

Fabio L. Spizzichino
Rome University La Sapienza, Rome, Italy

For m non-negative random variables $X_1,..,X_m$ consider the quantities

$$\alpha_j(A) = P(X_j = min_i X_i)$$

for any j ∈{1,2,...,m} and any subset A of {1,2,...,m}. Namely $\alpha_j(A)$ is the probability that X_j takes on the minimum value among all the other variables X_i, for i ∈A. When A exactly contains two elements, A:={i,j} say, the inequality $\alpha_j(A) \geq 1/2$ translates the condition that X_j stochastically precedes X_i. A multivariate concept of stochastic precedence can thus be considered by looking, for each subset A, at the element (or the elements) j such that $\alpha_j(A) \geq \alpha_i(A)$ for all i ∈A. Let us denote by $X_{1:m}, \dots, X_{m:m}$ the order statistics corresponding to the variables $X_1,...,X_m$, and assume the no-tie condition $P(X_j = X_i) = 0$, for i≠j. Thus we can consider the random indices $J_1, ..., J_m$ defined through the position $J_r = i \Leftrightarrow X_{r:m} = X_i$. The joint distribution of $(J_1, ..., J_m)$ is then a probability distribution over the set of the permutations of the elements $(J_1, ..., J_m)$ Such distribution will be denoted by Π_J.

In the first part of the talk, attention will be mainly concentrated on some basic properties of the family $\{\alpha_j(A), A \subseteq \{1,2,\dots,m\}, j \in A\}$

In particular we will analyze the relations between A and the distribution Π_J. We will furthermore point out an interesting role of the special class of survival models of the type time-homogeneous load-sharing, which are suggested from the context of reliability theory. It will be shown that such models can be used in the construction of specific distributions for $(X_1,..., X_m)$, corresponding to an arbitrary choice of Π_J.

In the second part of the talk it will be highlighted that, in the analysis of A one can encounter different types of paradoxes, which are similar to the paradoxes emerging in the frame of voting theory. Within such a theory, the elements 1,...,m, assume the meaning of candidates in different possible elections. In particular, the possibility of non-transitive behavior is a relevant aspect of stochastic precedence. This aspect can be seen as a phenomenon similar to the Condorcet Paradox of voting theory. Furthermore, other controversial situations can be observed when one inspects the probabilities $\alpha_j(A)$ for subsets A containing more than only two elements. Also for such situations that emerge in the analysis of the family A one can find strict similarity with paradoxes appearing in voting theory. In this respect we will clarify that, more than a similarity, a structural connection can actually be established between the two different fields of voting theory and analysis of the families of type A. On this purpose, we will first recall the concept of voting situation, as it emerges in a standard context of voting theory. We will then point out that voting situations are frequency distributions over the set of the permutations of the elements of {1,2,...,m}, whose role in that context is analogous to the one of the probability distributions Π_J in the study of A. This is joint work with Emilio De Santis at Rome University La Sapienza.

Invited and Contributed Talks

A Topological Clustering of Variables

Rafik Abdesselam
University of Lyon, Lumi`ere Lyon 2, ERIC - COACTIS Laboratories
Department of Economics and Management, 69365 Lyon, France

Clustering is one of the most important approaches to exploring multivariate data. The two most common unsupervised clustering strategies are hierarchical ascending clustering (HAC) and k-means partitioning used to identify groups of similar objects in a dataset to divide it in several homogeneous groups.The proposed topological clustering of variables, called TCV, studies an homogeneous set of variables defined on the same set of individuals, based on the notion of neighborhood graphs, some of these variables are more-or-less correlated or linked according to the type quantitative or qualitative of the variables. This topological data analysis approach can then be useful for dimension reduction and variable selection. Its a topological hierarchical clustering analysis of a set of variables which can be quantitative, qualitative or a mixture of both. It arranges variables into homogeneous groups according to their correlations or associations studied in a topological context of principal component analysis (PCA), multiple correspondence analysis (MCA) and mixed principal component analysis (MPCA). The proposed TCV is adapted to the type of data considered, its principle is presented

and illustrated using simple real datasets with quantitative, qualitative and mixed variables. The results of these illustrative examples are compared to those of other variables clustering approaches.

Keywords: Hierarchical clustering of variables, proximity measure, neighborhood graph, adjacency matrix, multivariate quantitative, qualitative and mixed data analysis, dimension reduction.

References:

1. Abdesselam, R.: Selection of proximity measures for a Topological Correspondence Analysis. Data Analysis and Applications 3, Computational, Classification, Financial, Statistical and Stochastic Methods, Vol.5, Part 2. Classification Data Analysis and Methods. Wiley, 103–120, 2020.
2. Batagelj, V., Bren, M.: Comparing resemblance measures. In Journal of classification, 12, 73–90, 1995.
3. Chavent M., Kuentz V., Liquet B., Saracco J., ClustOfVar: An R Package for the Clustering of Variables, Journal of Statistical Software. Vol. 50, 1-16, 2012.
4. Llobell, F. and Qannari, E.M.: Clustering datasets by means of CLUSTATIS with identification of atypical datasets. Application to sensometrics. Food Quality and Preference, Elsevier, 75, 97–104, 2019.
5. Saporta, G.: Simultaneous treatment of quantitative and qualitative data. In Attidela XXXV Riunone scientifica; Societa` Italiana di Statistica, 63–72, 1990.
6. Vigneau, E., Qannari, E.M.: Classification de variables around latent components. Communications in statistics Simulation and Computation, 32(4), 1131-1150, 2003.

Queueing Systems with Regenerative Input Flow

Larisa Afanaseva[1], Elena Bashtova[2]

[1]Department of Mathematics and Mechanics, Moscow State University, Moscow, Russia, [2]Department of Mathematics and Mechanics, Moscow State University, Moscow, Russia

We consider a multi-server queueing system in which primary customers arrive according to a regenerative flow. The system has m stochastically identical servers. Service times are independent identically distributed random variables and have an arbitrary distribution. An arriving customer finding one or more servers idle obtains service immediately. Customers

who find all servers busy go directly to the orbit from which repeat attempts to get into idle server. Two classes of such systems are considered. For the first class the rate of retrial requests depends on the number of customers on the orbit and for the second class the rate is a constant.
Based on the synchronization of the input flow and an auxiliary service process we establish the necessary and sufficient stability conditions for the models of both classes. Limit theorems for the number of customers on the orbit in the heavy traffic situation are established.
Approaches to statistical analysis of the system parameters when the number of customers on the orbit is observed are presented. Some applications are also given.
Keywords: Queueing systems, Retrial systems, Stability conditions, Limit theorems, Heavy traffic.
The research is supported by Russian Foundation for Basic Research, project 20-01-00487.

The Number of Completed Vacations for a Vacation Queueing System with Applications to Maintenance of Residential Buildings

Grigory Afanasyev
Moscow State University of Civil Engineering, Moscow, Russia

We consider a single-server queueing system with vacations and close-down that operates in the following manner. When the server returns from a vacation it observes the following rule. If there is at least one customer in the system, the server commences service. If the server finds the system empty a close-down period begins. If no customers have arrived during this period the server commences a vacation. Otherwise, the server begins service of the first arrived customer at the instant of this arrival. The input flow is supposed to be a Poison one outside of the vacation period and the flow of arrivals during vacation period has a single jump at the end of this period. Let η_n be the duration of the nth vacation period, T_n the moment of the nth vacation start and $\tau_n = T_{n+1} - T_n$, (n = 1,2,...). Define the random process $Y_n(t)$ as the number of customers which are present in the system at time $T_n + t$. The sequence $\{Y_n(t), t \leq$

$\eta_n\}$ consists of identically distributed independent random elements. We do not assume that $Y_n(t)$ is a Poison process with rate λ. It allows to study many new vacation queueing models. The proposed system is considered as the simplest model for the technical operation of residential buildings, which has two main functions for servicing the housing stock scheduled prophylactic inspection and repair of technical objects, as well as the elimination of sudden malfunctions of technical equipment. The main goal of this paper is analysis of the first mentioned problem. To do it we need to study the renewal process *N*(t) that is the number of completed vacations at time t. We calculate Laplace-Stieltjes transform $\varphi(s) = Ee^{-s\tau}$ and then obtain the formula for the function $(z,s) \quad = Ez^{N(t)}e^{-st}$. It allows to calculate moments $E\tau$ a $D\eta$ and prove the Central Limit Theorem for $N(t)$. Based on these results the management company may organize the maintenance of residential building.

Keywords: Queuing systems, Vacations, Close-down times, Residential buildings.

Research is supported by Russian Foundation for Basic Research, project 20-01-00487.

A different approach to current developments in the 21st century - Grouping European countries in terms of Mortality

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

[1,3]Department of Geography, Harokopio University, Athens, Greece,
[2]Department of Informatics and Telematics, Harokopio University, Athens,, Greece

This work describes human mortality through a different spatial approach. Analysis comprises twenty-two (22) European countries and examines their projected values of model parameters. The mortality model used is the Beta Gompertz Generalized Makeham distribution. The grouping of countries relies on two different methods: Pareto analysis and Cluster analysis; both are statistical techniques, applied in decision making, that

classify similar cases in respect to specific criteria. In this work grouping is based on the four parameters of the Beta Gompertz Generalized Makeham distribution, namely the infant mortality (θ), the aging - expressed as the number of older people in a country over 70 in the total population (ξ), the random risk factor depending on age (κ) and the random risk factor affecting the total (λ) of each country. Results help identifying some interesting similarities and differences across European countries.

Keywords: Gompertz and Makeham functions, BGGM distribution, mortality dynamics, projections.

Predicting event's counts in event-driven clinical trials accounting for cure and ongoing recruitment

Vladimir Anisimov

Data Science, Center for Design & Analysis, Amgen, 240 Cambridge Science Park, Cambridge, CB4 OWD, United Kingdom

We consider event-driven clinical trials where a sample size is defined by the number of some clinical events, e.g., oncology and cardiovascular trials. At the interim stage, one of the main tasks is predicting the number of particular events over time and time to reach specific milestones accounting also for the events that may happen for patients yet to be recruited in the future. Therefore, in the event-driven trials we need to model patient recruitment and the process of event's appearance all together. In the talk we developed a new approach to this problem which accounts for the opportunity of patients to be cured and also for dropout and lost-to-follow-up patients. The recruitment is modelled using a Poisson-gamma model developed in the previous publications. For the process of event's appearance, we assume that the time to the main event and the time to dropout are independent random variables and developed a few models using exponential, Weibull and log-normal distributions. Some results for the exponential model are published in [1]. This technique is supported by the developed software tools in R. The results are illustrated using some particular datasets.

References:

[1] V. Anisimov, Modern analytic techniques for predictive modeling of clinical trial operations, in book “Quantitative Methods in Pharmaceutical Research and Development: Concepts and Applications”, O. Marchenko, N. Katenka, Edt, Springer, 2020, 361-408.
Keywords: Predicting event’s counts, Patient recruitment, Event-driven clinical trial, Poisson-gamma recruitment model, Cure model, Dropout.

Application of the Neumann - Pearson method to assess the effectiveness of an expert

Valery Antonov[1], Victor Garbaruk[2], Victor Fomenko
[1]Peter the Great St. Petersburg Polytechnic University St. Petersburg, Russian Federation, [2]Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, Russian Federation

The paper provides an assessment of the effectiveness of an expert's work depending on his individual probabilistic characteristics. The case is considered when there is only one group of “k” experts in the council.
In the Neumann – Pearson algorithm, type 1 error can be set arbitrarily, and the algorithm allows minimize type 2 error for a selected value of type I error. Further, the parameters are selected from the condition that the type 1 error takes on a given value. Let each expert evaluate situations A and B as pA= 1- pB = p. For each k we find the value at which the probability of a complete error in making a collective decision takes a given value g = g (k). As k grows, the probability p will decrease monotonically. Then you can determine the relative efficiency of the expert's work depending on his individual indicator ─ the probability of correctly recognizing the situation A or B. The problem was analyzed as follows. For a fixed value of k, the individual probabilities p(k) of making a correct decision by each expert are calculated for different values of the number of the council k, in reality, k = 1,…8. Let us define the expert's labour efficiency as 1/k. As a result, you can get the dependence p(k) on 1/k.
Note that since the decision is made by the council on the basis of an unambiguously determined optimal algorithm, such an assessment of the expert's efficiency seems to be quite objective.
Keywords: Expert efficiency, Neumann - Pearson method, Individual probability.

Kendall Interaction Filter for Variable Screening in High Dimensional Classification

Youssef Anzarmou, Abdallah Mkhadri, Karim Oualkacha
University of Cadi Ayyad, Marrakech, Morocco

Accounting for important interaction effects can improve prediction of many statistical learning models. Identification of relevant interactions, however, is a challenging issue owing to their ultrahigh-dimensional nature. Interaction screening strategies can alleviate such issues. However, due to heavier tail distribution and complex dependence structure of interaction effects, innovative robust and/or model-free methods for screening interactions are required to better scale analysis of complex and high-throughput data. In this work, we develop a new model-free interaction screening method, termed Kendall Interaction Filter (KIF), for the classification in high-dimensional settings. KIF method suggests a weighted-sum measure, which compares the overall to the within-cluster Kendall's tau of pairs of predictors, to select interactive couples of features. The proposed KIF measure captures relevant interactions for the clusters response-variable, handles continuous, categorical or a mixture of continuous-categorical features, and is invariant under monotonic transformations. We show that KIF measure enjoys the sure screening property in the high-dimensional setting under mild conditions, without imposing sub-exponential moment assumptions on the features' distributions. We illustrate the favorable behavior of the proposed methodology compared to the methods in the same category using simulation studies, and we conduct real data analyses to demonstrate its utility.

Keywords: Interaction screening; Dimension reduction; Classification; Features ranking.

A Discrete Time Series Model for Covid-19

Anastasios Apsemidis[1], Nikolaos Demiris[2]
[1,2]Department of Statistics, School of Information Sciences and Technology, Athens University of Economics and Business, 76, Athens, Greece

Covid-19 is undoubtedly a pandemic that people will remember and one of the biggest events of the 21st century that has costed many lives and has had a large social impact and economic consequences. The deadly disease raised much scientific interest from a statistician's perspective and different models are proposed in the literature. The technique discussed here is a discrete time non-linear model based on daily data about deaths and confirmed cases in Greece, which follows the formulation of the standard SIR and SEIR epidemic models. Utilizing a suitable bivariate likelihood, the proposed model is capable of not only predicting the new deaths and the well-known index Rt, but the unobserved cases as well, which is generally a difficult task. To this end, mobility data obtained from Google and Apple are synthesized and analyzed, in order to infer the infection rate. Finally, suitable priors obtained from the literature are utilized in different parts of the model, leading to a large synthesis of information, so that this novel virus can be explained.
Keywords: Covid-19, SARS-CoV-2, discrete time series, nonlinear models, Bayesian.

Robust Methodology in Sufficient Dimension Reduction

Andreas Artemiou
School of Mathematics, Cardiff University, Cardiff, UK

Sufficient Dimension Reduction (SDR) is a supervised dimension reduction framework to address dimension reduction mainly in regression and classification settings. Classic methods in SDR involve the use of inverse moments. In this work we will discuss new methods to address the outliers in SDR methodology by using multivariate medians.
Keywords: Sufficient Dimension Reduction, Inverse Moments, Outliers, Multivariate Medians.

A Novel Approach to Multicollinearity Detection via Coefficient Penalization

Andreas Artemiou[1], Alex Karagrigoriou[2], Kimon Ntotsis[2]
[1]School of Mathematics, Cardiff University, Wales, UK, [2]Laboratory of Statistics and Data Analysis, University of the Aegean, Greece

In this new era, of Big Data Analytics, researchers must be able to handle more complex information than ever and produce highly validated results. With the arrival of this type of data, new obligations in statistical analysis surfaced. Many well established techniques require modifications in order to properly perform due to the involvement of new factors. When it comes to features interpretation, multicollinearity is the biggest issue that the researcher has to overcome. Extreme multicollinearity among them can prevent or mislead their proper interpretation, therefore various diagnostics have been established, e.g. Variance Inflation Factor, Condition Indices, Eigensystem Analysis, etc. for identification purposes. In this presentation a novel approach to multicollinearity detection occurs with the foundation of a new criterion for both multicollinearity detection and elimination. The pivotal characteristic of this criterion is that it manages to descry the features that have been configured based on other(s) of the same given dataset, through an aggregation of L1 and L2 regularizations, something that already existing criteria are unable to achieve and commonly indicate all involved features as multicollinear. With this criterion better model formation and estimation can be established.
Keywords: Multicollinearity Detection, Elastic Net Reguralization, Model Selection Criterion, Coefficient Penalization.

A new regression model for count compositions

Roberto Ascari[1], Sonia Migliorati[1]
[1]Department of Economics, Management and Statistics (DEMS), Università degli studi di Milano-Bicocca, Milano, Italy

Count compositions are vectors of non-negative integers summing to a fixed constant. The most popular distribution for this kind of data is the multinomial one, which has many advantages but is poorly parameterized in terms of the covariance matrix. An interesting approach to overcome this issue is to compound the multinomial distribution with a distribution defined on the simplex. For example, compounding the multinomial distribution with the Dirichlet leads to the well-known Dirichlet-multinomial (DM). With an additional parameter, the DM distribution fits real data better than the multinomial one, but its covariance structure may still be too rigid. The aim of this work is to propose a new distribution for count compositions and to develop a regression model based on it. The new distribution is obtained by compounding the multinomial with the extended flexible Dirichlet (Ongaro et al, 2020) and it can be expressed as a structured finite mixture with particular DM components. Thanks to this mixture structure and the additional parameters introduced, the new distribution can provide a better fit and an interesting interpretation in terms of latent groups. We compare the regression models based on these distributions through a simulation study. Inferential issues are dealt with a Bayesian approach through the Hamiltonian Monte Carlo algorithm.
Keywords: Multinomial data, Multivariate regression model, Bayesian, Mixture model, Compound distribution.

Prediction and analysis of the Extreme and Records values of air pollution data in Bekaa valley in Lebanon

Alya Atoui[1,2], Zaher Khraibani[2,3], Regis Moileron[1], Kamal Slim[4], Samir Abbad Andalousi[1]

[1]Laboratoire Eau Environnement et Systèmes Urbains (LEESU), Université Paris-Est Créteil, France, [2]Lebanese University, Rammal Rammal Laboratoy, phyToxE group, Nabatieh, Lebanon, [3]Department of Applied Mathematics, Lebanese University, Faculty of Sciences, Hadat, Lebanon, [4]Lebanese Atomic Energy Commission, National Council for Scientific Research (CNRS), Beirut, Lebanon

During the last decades, extreme events modeling has been of particular interest in various fields related to earth and environmental sciences, such as; hydrology, seismology, and applications related to climate change or air pollution monitoring. Extreme Value Theory (EVT) deals with the modeling of extreme events where the main purpose is the estimation of

an extreme quantile and the return level associated with its return period using the two classical approaches: GEV and POT. Given daily measurements of O3, CO, NO, NO2, PM10, and PM2.5 from 2013-2019 at Zahle station in the Bekaa region of Lebanon, we predict the return level of the concentrations of pollutants over several years (20,30,50 years) to create an alert system for the future. Then, we compare the different estimation methods of quantiles, maximum likelihood, and methods of weighted moments. The comparison is based essentially on the calculation of the bias and RMSE of each estimator, from Monte Carlo realizations of the initial model. After determining the extreme distribution of the pollutants concentrations we use a new approach to predict future records based on record theory. Thus we calculate an interval of pollutant concentration prediction. Finally, we evaluate the quality of prediction of the measurements by choosing the Best Linear Unbiased Predictor.
Keywords: Air pollution, Extreme, Records, Prediction, Return level, Lebanon.

Cluster analysis after multiple imputation

Vincent Audigier[1], Ndèye Niang[2]
[1,2]CNAM, Laboratoire Cedric-MSDMA, Paris, France

Clustering individuals is an essential task for data science. However, data are often incomplete and clustering algorithms cannot be directly applied on incomplete data. A popular way to deal with missing values in data analysis consists in using multiple imputation. Multiple imputation (MI) consists of 3 steps: 1) the imputation of the data set according to an imputation model 2) the analysis of each imputed data set according to an analysis model 3) the pooling of the analysis results according to the Rubin's rules. Such methods are mainly used for inference in linear models, but not for cluster analysis. One major issue for applying cluster analysis after MI is how to apply an equivalent of Rubin's rules in this context, i.e. how to pool the partitions obtained from each imputed data set and how to assess the instability of the clustering with missing values? Based on theoretical arguments, we first propose to use median partition-based methods for partitions pooling. In particular, simulations show non-negative matrix factorization methods are theoretically and computationally attractive for achieving this goal. Then, based on

bootstrap theory, we additionally propose a new rule for assessing the cluster stability with missing values. From a practical point of view, our first rule improves partition accuracy with missing values. We highlight the accuracy is sensitive to the number of imputed data sets (M), particularly when the number of individuals is small or when the proportion of missing values is large. Simulations show M=50 is generally enough. The second rule allows calculation of an additional between instability related to missing values. We show it provides a new way for dealing with the number of clusters when data are incomplete, which can be particularly useful for distancebased clustering methods. Furthermore, it allows a sensitivity analysis of the partition to the imputation model.

Keywords: Clustering; Uncertainty; Missing Data; Multiple imputation; Consensus clustering; Rubin's rules.

Recent Developments in Conditional Frontier Analysis

Luiza Bădin[1,2]

[1]Bucharest University of Economic Studies, [2]Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Bucharest, Romania

Recent theoretical and empirical studies on conditional frontier models highlight the importance of conditional efficiency measures as a general nonparametric way to treat appropriately the presence of environmental factors in a production process. Conditional efficiency measures are based on the idea that the production process can be described as being conditioned by given values of the external, environmental factors. These factors can be included in the frontier model as exogenous variables and can help explaining the efficiency differentials and improving the managerial policy of the evaluated units. Nonparametric conditional frontier models allow for a complete and general handling of heterogeneity, without relying on strong parametric or semi-parametric assumptions, many times unrealistic. An important aspect is the explanation of differences in the efficiency levels achieved by economic producers that are facing different environmental and external conditions.

In this paper we explore the most recent methodological developments in conditional frontier analysis, providing useful insights on their practical implementation.

Keywords: Nonparametric frontier models, conditional efficiency, external factors, bandwidth, bootstrap.

Acknowledgment: This work was supported by a grant of the Romanian Ministery of Education and Research, CNCS-UEFISCDI, project number PN-III-P4-ID-PCE-2020-1112, within PNCDI III.

Structure of the Particle Field in Branching Walks with Generation Centers of Particles at Every Lattice Point

Daria Balashova

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

We consider a continuous-time symmetric branching random walk on a multidimensional lattice. Initial population consists of a single particle at each point on the lattice. We assume that reproduction law is described by a Markov branching process at every lattice point with the possibility of branching into an arbitrary number of particles. We study the structure of the particle subpopulation generated by the initial particle situated at some lattice point. Equations of the first and second factorial moments of the subpopulation size are obtained. The areas of “intermittency” are found in the case of the supercritical law of branching of particles and superexponential light tails of underlying random walk. In case of critical branching into two offspring, asymptotic behavior of conditional second moment was found for random walks with superexponential light or heavy tails. The convergence of the distribution of the particle field to the stationary one is proved for processes with a recurrent walk and a critical Markov branching process at every lattice point.

The research was supported by the Russian Foundation for the Basic Research (RFBR), project No. 20-01-00487.

Keywords: Branching random walk, Multidimensional lattices, Limit theorem, Intermittency.

Log-linear models in the presence of out-of-scope units

Veronica Ballerini[1], Li-Chun Zhang[2]
[1]Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Rome, [2]Department of Social Statistics and Demography, University of Southampton; Statistics Norway and Department of Mathematics, University of Oslo

For the last few years, many National Statistics Institutes have been shifting from a census-based statistics paradigm to a register-based one. The final goal consists of letting the sample surveys support the information collected by the administrative registers and not the other way around. Nevertheless, several methodological issues emerge. Particular attention is devoted to the overcoverage, i.e., the possibility that one or more data sources erroneously include some out-of-scope units. In this work, we present an alternative modelisation of the latent erroneous counts in a capture-recapture framework, simultaneously addressing both the overcoverage and undercoverage problem. We rely on the Pseudo Conditional Independence assumption (Zhang 2019) to include non-negligible prior information. Moreover, we address model selection, which is not trivial in the framework of log-linear models when there are a few (or even zero) degrees of freedom.
Keywords: capture-recapture, Bayesian log-linear models, erroneous enumerations, latent variables, model selection.

Detecting change-points in multivariate hydrological time series

Dominika Ballová
Faculty of Civil Engineering, Slovak University of Technology, Bratislava

The detection of change-points is crucial in various fields of study. In hydrological processes the detection of changes might help predict extreme events like floods and drouth in the future development of the time series. Sudden switch in parameters, mostly in the mean value of

the distribution is represented by a change-point. In this paper we focus on changes in the development of multivariate hydrological time series. Since the data follow non normal distribution, we obtained our results by means of nonparametric statistical methods. First goal is to detect the change-points for each univariate time series separately, then we seek to find the multivariate change-points and compare the results. In the case of multivariate time series we seek common change-points of the series using several statistical techniques as correlation analysis or copulas. These methods are applied and compared both on simulated data and on hydrological processes measured on some of the main Slovak rivers.

Keywords: Change-points, Hydrological Processes, Time series.
Acknowledgement. This work was supported by Slovak Research and Development Agency under contract No.APVV-17-0066 and the grant No. APVV-18-0052.

A monotone single index model for missing-at-random longitudinal proportion data

Dipankar Bandyopadhyayi
Virginia Commonwealth University, Richmond, United States

Beta distributions are commonly used to model proportion valued response variables, commonly encountered in longitudinal studies. In this article, we develop semi-parametric Beta regression models for proportion valued responses, where the aggregate covariate effect is summarized, and flexibly modeled using an interpretable monotone time-varying single index transform of a linear combination of the potential covariates. We utilize the potential of single index models, which are effective dimension reduction tools, and accommodates link function misspecification in generalized linear mixed models. Our Bayesian methodology incorporates missing-at-random feature of the proportion response, and utilize Hamiltonian Monte Carlo sampling to conduct inference. We explore finite-sample frequentist properties of our estimates, and assess the robustness via detailed simulation studies. Finally, we illustrate our methodology via application to a motivating longitudinal dataset on obesity research, recording proportion body fat.

A Goodness of Fit Test based on Directed Divergence Measures

Vlad Stefan Barbu[1], Thomas Gkelsinis[2], Alex Karagrigoriou[3]
[1,2]Laboratory of Mathematics Raphaël Salem, University of RouenNormandy, France, 3Laboratory of Statistics and Data Analysis, University of the Aegean, Greece

In this paper we present a Directed Goodness of Fit (DGoF) test associated with a directed test of homogeneity between two samples and we study their performance. These types of tests are constructed based on a particular type of discrepancy measures called directed divergences. These measures allow us to focus on specific subsets of the support without, at the same time, loosing the information of the others. With this method we achieve a significantly more sensitive test as compared to the classical ones with comparable error rates.
Keywords: Goodness of Fit, Divergence Measures, Directed Information Measures, Homogeneity test.

Statistical Inference for Multi State Systems: The Kumaraswamy case

Vlad Stefan Barbu[1], Alex Karagrigoriou[2], Andreas Makrides[2]
[1]Raphaël Salem, UMR 6085, Avenue de l'Universite, Saint-'Etienne-du-Rouvray, France, [2]Department of Mathematics, University of the Aegean, Samos, Greece

Semi-Markov processes are typical tools for modeling multi state systems by allowing several distributions for sojourn times. In this work, the sojourn times are considered to follow the Kumaraswamy distribution with two shape parameters which belongs to a general class of distributions with the advantage that is closed under minima. We provide estimators for the parameters of interest and we evaluate our methodology using an exhaustive simulation study.
Keywords: Multi-state system, Semi-Markov processes, Kumaraswamy distribution, Beta distribution, Reliability analysis, Parameter estimation.

Moment Based Approximation for a Semi Markovian Inventory Model of Type (s,S)

Aslı Bektaş Kamışlık[1], Feyrouz Baghezza[2], Tülay Kesemen[3], Tahir Khaniyev[4]

[1]Recep Tayyip Erdoğan University Fener Mahallesi, Merkez/Rize, Turkey, [2,3]Karadeniz Technical University, Üniversite, Ortahisar/Trabzon, Turkey, [4]TOBB Economy and Technology University, Çankaya/Ankara, Turkey

In this study an inventory model of type (s,S) is considered and constructed by using a semi-Markovian renewal reward process. When examining the characteristics of semi-Markovian stock control models, the renewal function generated by the demand random variables plays an important role. All of the renewal functions proposed in the literature so far require the functional form of the distribution function of demand random variables F (t). This study is based on the approximation formula proposed by Nirmal S. Kamboo et. al (2012) for the renewal function U(t) generated by demand random variables. In this method, if the first three moments of the demand random variables are finite and known there is no need to know the explicit formula of the distribution function F(t). It is relevant for demand distributions with coefficient of variation smaller than one. Using this method we were able to obtain seven term asymptotic expansion for ergodic distribution of the considered process. Our approximation is applicable when demand random variables have certain important distributions like exponential, mixture of two exponentials, Erlang with two phases and K2 (Coxian-2) used widely in application.

Keywords: Moment based approximation, Semi-Markovian stochastic process, Phase type distributions, Coxian-2 distribution.

Statistical analysis of the non-ergodic fractional Ornstein-Uhlenbeck process with periodic mean

Rachid Belfadli[1], Khalifa Es-Sebaiy[2], Fatima-Ezzahra Farah[3]
[1]Research Group of Geometry, Stochastic Analysis and Applications, Department of Mathematics, Faculty of Sciences and Techniques, Cadi Ayyad University, Marrakech, Morocco, [2]Department of Mathematics, Faculty of Science, Kuwait University, Kuwait, [3]National School of Applied Sciences-Marrakech, Cadi Ayyad University, Marrakech, Morocco

Consider a periodic, mean-reverting Ornstein-Uhlenbeck process $X = \{Xt, t \geq 0\}$ of the form $dX_t = (L(t) + \alpha X_t)dt + dB_t^H$, $t \geq 0$, where $L(t) = \sum_{i=1}^{p} \mu_i \phi_i(t)$ is a periodic parametric function, and $\{BtH, t \geq 0\}$ is a fractional Brownian motion of Hurst parameter $\frac{1}{2} \leq H < 1$. In the ergodic case $\alpha < 0$, the parametric estimation of $(\mu_1,...,\mu_p,\alpha)$ based on continuous-time observation of X has been considered in Dehling et al. [?], and in Dehling et al. for $H = \frac{1}{2}$, and $\frac{1}{2} < H < 1$, respectively. In this paper we consider the non-ergodic $\alpha > 0$ case, and for all $\frac{1}{2} \leq H < 1$. We analyze the strong consistency and the asymptotic distribution for the estimator of $(\mu_1,...,\mu_p,\alpha)$ when the whole trajectory of X is observed.

Keywords: Parameter estimation; Strong consistency; Joint asymptotic distribution; Fractional Ornstein-Uhlenbeck process; Periodic mean function; Young integral.

Stochastic Models for Nonlinear Systems of PDEs

Yana Belopolskaya
Department of Mathematics, SPbGASU, St.Petersburg, Russia

A number of systems of nonlinear PDEs arising in physics, chemistry, biology and other fields admit an interpretation as systems of forward Kolmogorov equations for certain stochastic processes. This interpretation leads to better understanding of phenomena described by these systems and besides allows to develop new numerical algorithms to construct approximate solutions to initial-boundary value problems for

these systems. In this talk we consider several examples of probabilistic approach to solutions of the Cauchy problem for systems of nonlinear PDEs. In particular, we consider the Cauchy problem for the MHD-Burgers system arising in plasma physics, for some systems describing chemotaxis processes and prey-predator type systems arising in population dynamics. Within a probabilistic approach to solution of the Cauchy problem for a system of nonlinear PDEs we derive stochastic differential equations (SDEs) for stochastic processes to be used to construct probabilistic representations for solutions of original problems. Since original problems are nonlinear, we obtain SDEs with coefficients depending on unknown functions and we have to find the corresponding closing relations. As a result, we obtain closed systems of stochastic equations to be solved and investigated. Finally, for each of the derived SDE systems we verify that its solution defines a solution of the considered Cauchy problem for the original PDE system.

Keywords: Stochastic process, Systems of nonlinear PDEs, Cauchy problem.

Bandwidth selection for M-estimation under truncated, censored and dependent data

Hassiba Benseradj[1], Zohra Guessoum[2]

[1]Faculty of Sciences, University of Boumerdes UMBB, Algeria, [2]Lab. MSTD, Faculty of Mathematics, Algeria

This paper is concerned with the problem of selecting a suitable bandwidth, for the M-estimator of the robust regression function from left truncated and right censored data (LTRC), under strong mixing condition: We provide an asymptotic expression for the mean integrated squared error (MISE) of this estimator. As a consequence, a bandwidth selector based on iterative plug-in ideas is introduced. We also present a robust version of the Least Square Cross-Validation (RLSCV) bandwidth selection. A simulation study is investigated to examine the practical performance of both two methods.

Keywords: Alpha-mixing; Bandwidth selection; M-estimator.

Pricing energy quanto options: a regime-switching framework with stochastic interest rates

Fred Espen Benth, Griselda Deelstra, Sinem Kozpınar
Ankara, Turkey

This paper investigates the valuation of energy quanto options when the underlying price processes are modelled by Markov-modulated additive processes in the presence of regime-switching stochastic interest rates. Using change of measure and fast Fourier transform techniques, we obtain a pricing formula for energy quanto options on futures. We show the precision of our pricing formula on a quanto option written on temperature and electricity future prices.
(Speaker: Sinem Kozpınar)
Keywords: Energy quanto option, Markov-modulated additive processes, stochastic interest rates, futures prices.

Discussion over Sequential Multi-block Methods

Alessandra Biancolillo
Department of Physical and Chemical Sciences, University of L'Aquila, Coppito, L'Aquila, Italy

Data fusion strategies represent powerful tools for handling multi-block data sets. Sequential multi-block methods provide benefits and peculiarities which could be particularly useful in various situations. Among the others, Sequential and Orthogonalized Partial Least Squares (SO-PLS) and Sequential and Orthogonalized Covariance Selection (SOCovSel provide relevant benefits from both the predictive and the interpretative point of view. In the present work, in order to highlight the advantages provided by these approaches, few examples and applications of these methods on real data are discussed.
Keywords: Multi-Block Methods, Data Fusion, Regression, Classification.

Exploring chronic diseases' spatial patterns: thyroid cancer in Sicilian volcanic areas

Francesca Bitonti[1], Angelo Mazza[2]
[1,2]Department of Economics, University of Catania, Catania, Italy

Whereas spatial analyses of infectious diseases have a long tradition, only recently, a consistent interest emerged towards the geography of chronic diseases and their environmental correlations. In this work, we explore spatial variations in the prevalence of thyroid cancer, taking into account the demographic heterogeneity in the at-risk population at the small-area level. This work aims at enhancing the existing research about thyroid incidence in volcanic areas by analyzing spatial patterns of thyroid cancer cases in Mt Etna's area, in the eastern part of Sicily. It is known from the medical literature that several constituents of volcanic lava and ashes, such as radioactive and heavy metals, are involved in the pathogenesis of thyroid cancer via the biocontamination of atmosphere, soil, and aquifers. Here, we exploit a unique dataset that allowed us to geocode the geographic location of cases at the household level, whereas all studies that we are aware of use aggregated data. Applying the generalized Ripley's K-function as a means for detecting spatial clustering in inhomogeneous point patterns, we aimed to disentangle the spatial aggregation of thyroid cancer cases due to the proximity to a volcanic area from that due to the geographic variations in the density of the population at risk and other concomitant environmental risk factors.
Our preliminary findings seem to confirm a vast empirical literature that has revealed an increased thyroid cancer incidence in volcanic areas, such as Island, Hawaii, and the Philippines, where an intense basaltic volcanic activity has also been long detected; furthermore, parts of the Etna volcanic area seem to be more affected than others.
Keywords: spatial analysis, thyroid cancer, point processes, inhomogeneous Ripley's K-function, volcanic area, spatial clustering.

Random orthogonal projections based estimators for nonparametric and linear functional regressions

Bilel Bousselmi[1], Jean-François Dupuy[2], Abderrazek Karoui[3]

[1,2]University of Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, Rennes, France, [3]University of Carthage,Department of Mathematics, Faculty of Sciences of Bizerte, Tunisia

The aim of this work is two fold. Firstly, we develop a Jacobi polynomials based scheme for the stable solution of non parametric regression problems. Secondly, we develop an orthogonal projection based estimator of the stable solution of linear functional regression (LFR) problem. More precisely, the first sceme provides us with a two parameters family of nonparametric regression estimators. These estimators are constructed by using a random kernel associated with a finite orthonormal set of Jacobi polynomials $\{P_k^{(\alpha,\beta)}, 0 \leq k \leq N\}$ where $\alpha, \beta \geq -\frac{1}{2}.$ A convergence analysis of these two parameters estimators is done under the assumption that the true regression function has some Lipschitz regularity or it is the restriction to $I=\left[-1,1\right]$ of bandlimited function. Moreover, we show that our random orthogonal projection based LFR estimator for the slope function is stable. This is done by providing a convenient upper bound for the 2-condition number of this random pseudo-inverse matrix. Finally, we illustrate the different results of this work by some numerical simulations

Keywords: Nonparametric regression, empirical projection, Jacobi kernel, linear functional regression, random pseudo-inverse, condition number.

Analysis of Blockchain Based Databases in Web Applications

Orhun Ceng Bozo[1], Rüya Şamlı[2]
[1]Istanbul University Cerrahpaşa, İstanbul, Turkey, [2]Department of Computer Engineering, İstanbul University Cerrahpaşa, İstanbul, Turkey

Blockchain; since it is a technology used for creating, registering and sharing data, is used as an alternative to existing databases or together with existing database technologies. In this study, the functions of relational, non-relational databases and blockchain-based databases in web applications were compared. Whether these systems with different capabilities have performance differences reflected on users, whether security vulnerabilities while meeting the user needs, and advantages and disadvantages of using blockchain were evaluated. The types of blockchain technologies and the use of different combinations of database types in different situations have been analyzed in terms of functionality, security, performance in web applications. Significant differences were noted in the results, especially in which steps.
Keywords: Database, Sql, Nosql, Blockchain, Blockchain Based Database, Blockchain in Web Applications.

Multiblock prediction model through Partial Least Squares Path Modeling

Stéphanie Bougeard[1], Pasquale Dolce[2], Véronique Cariou[3], Mohamed Hanafi[3]
[1]Department of Epidemiology, Anses, Ploufragan, France, [2]Department of Public Health, University of Naples Federico II, Italy, [3]StatSC, Oniris, INRAe, Nantes, France

An important issue in modern science is the integration of multiple data sources, considered as datasets partitioned into blocks of variables on the same set of observations. In some situations, relationships between blocks can be set-up from prior knowledge leading to a path modeling structure. In this scope, several methods were proposed, such as PLS Path Modeling (PLS-PM) [Wold, 1982], regularized Generalized

Structured Component Analysis [Hwang, 2004], regularized Generalized Canonical Correlation Analysis [Tenenhaus, 2011], THEmatic Equation Model Exploration [Bry, 2012] or Path-ComDim [Cariou et al., 2018] among others. In most of these approaches, the predictive purpose has been often neglected while it appears of paramount interest for: (i) evaluating and comparing method performances (ii) selecting the optimal number of block-components with cross-validation, and (iii) predicting new observations. Herein, an explicit formulation of a prediction model in the context of a path structure is proposed, which can also handle blocks multidimensionality. This model is applied to PLS Path Modeling this method being the most well-known one but its versatility provides an adaptation to any other multiblock methods. Two different estimations are proposed, both being related to the structural model. To take advantage of the full multidimensional potential of blocks, multiple block-components are taken into account. A deflation strategy based on the residuals of the model is proposed. The model and its two estimations are evaluated on the basis of real datasets involving two and three blocks.
Keywords: Structural equation; path-modeling; multiblock methods; component methods; prediction model; PLS regression.

Generalized linear regression with multiple explanatory arrays: the Supervised Component methodology

Xavier Bry[1], Catherine Trottier[1,2]

[1]IMAG - Université de Montpellier, France, [2]Université Paul Valéry, Montpellier, France

The number of explanatory variables (EV) in statistical models increasing dramatically, generalized linear modeling (GLM) is currently faced with two problems. The first is that the high dimension and redundancy of EVs cause it to break down, unless some regularization is performed. The second is a growing difficulty in interpreting the results: even though the estimation may yield a satisfying prediction formula of the responses, it often fails to shed enough light on the underlying explanatory dimensions. The most widespread way to regularize a GLM is combining its likelihood with a penalty on some norm of the coefficient-vector, as do ridge, LASSO, and elastic-net regressions. We propose a PLS-type alternative basing the linear predictor on components that use the redundancy of EVs to be both

more reliable and interpretable. We combine the likelihood of the GLM with a criterion measuring the "structural relevance" of components. The combined criterion is maximized under a unit-norm constraint on the component's loading vector. A sequence of uncorrelated components is extracted, from which the regularized linear predictor is obtained. The latter being decomposed on the explanatory components, it is all the more interpretable. The best number of components is selected through cross-validation. We extend this supervised-component methodology to situations where the EVs are partitioned into subsets, called themes, each having some conceptual unity. Doing so allows to extract components that are conceptually specialized, and better separate the effects of the themes on the responses. We show how to determine the number of useful components in each theme, and to calculate confidence intervals on the coefficients of EVs through bootstrapping. The methodology is applied to a Poisson-regression model of the abundances of tree species in the Congo basin rainforest, and to Coxregression, in order to analyze the entrance of Senegalese men into polygamy.

Keywords: Components, GLM, regression, regularization, SCGLR.

Optimization and Asymptotic Analysis of Insurance Models

Ekaterina Bulinskaya
Lomonosov Moscow State University, Moscow, Russia

Insurance is the oldest domain of applied probability, see [1]. Moreover, the mathematical models arising there can be used in other areas of applied probability. So, optimization of insurance models' performance and their asymptotic analysis are very important. Modern period in actuarial sciences is characterized by investigation of complex systems and employment of sophisticated mathematical tools. Discrete-time models became popular, since in many cases they describe more precisely the real situation. They can also serve as approximation of corresponding continuous-time models, see [4]. Hence, we study several discrete- and continuous-time models in the framework of reliability and cost approaches. Reinsurance, dividends, bank loans and investments are controls in optimization problems, see [2]. Stochastic orders of random

variables provide the possibility to compare various models. Parameters estimation and models' stability with respect to small fluctuations of parameters and perturbations of underlying distributions are treated as well. For this purpose, we use the methods of global and local sensitivity analysis and probability metrics, see [3].

The research is partially supported by the Russian Foundation for Basic Research, project 20-01-00487.

References:

[1] Bulinskaya E. (2017) New research directions in modern actuarial sciences. In: Springer Proc. in Math. and Statist. 208, pp. 349—408.

[2] Bulinskaya E. (2018) Asymptotic analysis and optimization of some insurance models. ASMBI 34 (6): 762—773.

[3] Bulinskaya E., Gusak J. (2016) Optimal control and sensitivity analysis for two risk models. Communications in Statistics -- Simulation and Computation 45: 1451—1466.

[4] Dickson D.C.M, Waters H. (2004) Some optimal dividends problems. ASTIN Bulletin, 34: 49—74.

Keywords: Optimization, Insurance Models, Asymptotic Analysis, Dividends, Reinsurance.

On Fluctuations of the Propagation Front of a Catalytic Branching Random Walk

Ekaterina Vl. Bulinskaya

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

Recall that the model of a catalytic branching random walk (CBRW) on an integer lattice combines two mechanisms: particles movement and their splitting in the presence of catalysts located at fixed points of the lattice.

In [1] the authors study the maximum Mt of a supercritical CBRW on an integer line at time t. They prove that Mt grows almost surely linearly with a certain rate μ and also analyze the fluctuations Mt – μt, as time tends to infinity. We discuss extension of the results in [1] to the case of multidimensional lattice. In contrast to the one-dimensional case, instead of Mt we consider the propagation front of the particle population in CBRW. Moreover, our approach covers models with an arbitrary finite number of catalysts, whereas [1] mostly allows only the single catalyst. In this regard we employ the results published in papers [2] and [3]. Beside

the asymptotic analysis of solution to the derived system of non-linear integral equations, the proofs involve renewal theorems for systems of renewal equations, many-to-one formula, martingale change of measure, large deviation theory, convex analysis and the coupling method. Some new problems of the population evolution are tackled as well.

This research was carried out with the financial support of the Russian Science Foundation (grant 17-11-01173-Ext) at Novosibirsk State University.

References:

[1] Ph.Carmona, Y.Hu, The spread of a catalytic branching random walk, Ann. Inst. Henri Poincaré Probab. Stat., 50 (2014), pp. 327—351.

[2] E.Vl. Bulinskaya, Spread of a catalytic branching random walk on a multidimensional lattice, Stochastic Process. Appl., 128 (2018), pp. 2325—2340.

[3] E.Vl. Bulinskaya, Fluctuations of the propagation front of a catalytic branching walk, Theory Probab. Appl., 64 (2020), pp. 513—534.

Keywords: Branching random walk, Propagation front, Spread of population, Supercritical regime, Light tails.

Workshop description
Fundamental problems arising in the analysis of applied stochastic models of complex structure

Ekaterina Bulinskaya[1], Co-organizer Elena Yarovaya[2]

[1,2]Lomonosov Moscow State University, Moscow, Russia

The aim of all the presentations is to study the behavior of stochastic processes describing the evolution of systems of complex structure. Special attention will be paid to processes whose elements can give offsprings, die and move on space in homogeneous or non-homogeneous environments, as well as stochastic models with various control processes. The fundamental problems include development of new, previously unexplored, models of such processes, the study of their asymptotic behavior and the proof of new limit theorems, as well as the optimization of the behavior of a series of systems under conditions of incomplete information. The range of such processes is wide and

describes not only Markov, but also semi-Markov and non-Markovian models. The problems supposed to be solved within the framework of the project are of great theoretical importance for the development of new areas of probability theory and stochastic processes. At the same time investigated stochastic process ses allow to explain the effects that arise when studying the dynamics of various types of populations, queuing and insurance systems. The models introduced and methods developed will be useful for investigation of other areas of applied probability.

Feature Selection in Mixed Models

Alexander Bulinski
Steklov Mathematical Institute of Russian Academy of Sciences, Moscow

There are many important problems in economy, medicine, biology where one has to study multidimensional data for identification of the relevant (in a sense) factors having impact on a random response under consideration, see, e.g., [1]. In contrast to the correlation coefficient, one applies the basic information theory concepts to discover nonlinear links between random vectors. In this regard we mention the Shannon entropy, mutual information, conditional mutual information and f-divergences, e.g., the Kullback – Leibler one. We consider mixed models where some features form a vector having a density and others take values in discrete sets. Note that this model comprises the classical logistic regression. Statistical estimates of the mentioned characteristics, for instance mutual information, play essential role in the feature selection technique. We develop results of [2], [3] on principle asymptotic properties of such estimates involving k-nearest neighbor statistics. The problem of feature selection methods stability is treated as well. Moreover, we discuss the stopping criteria for employed algorithms. Theoretical results are supplemented by computer simulations. The work is supported by the Russian Science Foundation under grant 19-11-00290 and performed at the Steklov Mathematical Institute of Russian Academy of Sciences.

References:

[1] Bolon-Canedo V., Alonso-Betanzos A. Recent Advances in Ensembles for Feature Selection. Springer, 2018.

[2] Bulinski A., Kozhevin A. Statistical estimation of conditional Shennon entropy. ESAIM: PS 23 (2019), p. 350-386.
[3] Bulinski A., Kozhevin A. Statistical estimation of mutual information for mixed model. Methodology and Computing in Applied Probability, http://doi.org/10.1007/s11009-020-09-802-0. Published online 01 July 2020, p.1-20.
Keywords: Feature selection, Mixed model, Mutual information, Statistical estimates, Asymptotic properties.

Quantum Entropy Quantifiers of Non-Markovianity of Open Systems Dynamics

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

In contrast to memoryless evolution of open quantum systems modelled by analogy with classical Markov processes more realistic models reflect diverse quantum memory effects. Much attention has been paid recently to interpretation and classification of such non-Markovian dynamics; see [1] for discussion of main approaches to revealing memory revivals: via information backflow into the open system interacting with environment versus manifestations of dynamical maps non-divisibility. One also has to assess such effects quantitatively. Beside the proposed monitoring of system states distinguishability increase, as displayed by evolution of the trace distance, it is essential to develop quantum information methods to detect and quantify the non-Markovian features of dynamics. For such aim we study the employment of Quantum Jensen-Shannon divergence. As was recently established, its square root possesses all the properties of metric, likewise the counterpart used in classical information theory to quantify dissimilarity of probability measures. This new metric on the quantum states space has advantages over the trace distance. However, a drawback is that the properties of many quantum entropic characteristics were investigated in case of finite-dimensional Hilbert space underlying

the quantum system. Our goal is to extend that analysis to the important case of infinite-dimensional systems, using some methods of [2],[3].

References:

[1] H.-P.Breuer, E.-M.Laine, J.Piilo and B.Vaccini. Non-Markovian dynamics in open quantum systems. Rev. Mod. Phys., 2016, 88, 021002.

[2] M.E.Shirokov, A.V.Bulinski. Lower bounds on distances from a given quantum channel to certain classes of channels. J. of Mathematical Sciences, 2019, 241, No 2, 237-244.

[3] M.E.Shirokov, A.V.Bulinski. On quantum channels and operations preserving finiteness of the von Neumann entropy. Lobachevskii J. of Mathematics, 2020, 41, No 12, 2383-2396.

Keywords: Open quantum system, Non-Markovian evolution, Non-divisible maps, Information backflow, Distance bounds, Quantum relative entropy, Quantum Jensen-Shannon divergence.

Statistical Analysis of Traffic Volume in the 25 de Abril Bridge

Frederico Caeiro[1], Ayana Mateus[2], Conceição Veiga de Almeida[3]

[1,2]Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia & Centro de Matemática e Aplicações (CMA), 2829-516 Caparica, Portugal

[3]Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia & Millennium BCP, Portugal

Bridges are important structures. They are used on land transportation to connect different points that usually are inaccessible. Loading forces due to traffic flow are an important physical factor that affects the bridge structural reliability. Thus, for safety assessment it is important to monitor and study traffic bridge. In this work, we analyse the daily traffic data on the 25 de Abril Bridge in Portugal. The aim is to study the tail distribution.

Acknowledgments: This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações).

Keywords: Extreme Value Theory, Tail Distribution.

Implied Volatility under HJM-type Stochastic Volatility model

Thi Diu Cap[1], Ying Ni[1]
[1]Division of Mathematics and Physics, Ma¨lardalen University, Sweden

In this research, we propose a new and simple approach of extending the single-factor Heston stochastic volatility model to a more flexible one in solving option pricing problems. In this approach, the volatility process for the underlying asset dynamics depends on the time to maturity of the option. As this idea is inspired by the Heath-Jarrow-Morton framework which models the evolution of the full dynamics of forward rate curves for various maturities, we name this approach as the HJMtype stochastic volatility (HJM-SV) model. We conduct an empirical analysis by calibrating this model to real-market option data for underlying assets including an equity stock (ABB stock) and three market indices (EURO STOXX 50, S&P 500 and OMX index). For two of the aforementioned underlying assets, we consider two separated time spans of before and after the COVID-19 economic downturn and use one-year-long data respectively. We address the question as how well the new HJMSV type model captures the feature of implied volatility surface given by the market data.
Keywords: Implied volatility surface, stochastic volatility model, HJM framework.

Analysis of three-way data by means of rank (Lr,Lr,1) decomposition

Véronique Cariou[1], Philippe Courcoux[1], Mohamed Hanafi[1]
[1]StatSC, ONIRIS, INRAE, 44322 Nantes France

In a large variety of domains, the datasets under study correspond to a three-way array structure. For instance, in sensory evaluation, practitioners are often faced to a three-way structure defined as a (sample x attribute x assessor) array. Multivariate analyses dedicated to such structures fit under the umbrella of three-way (or more generally multiway)

analysis. Among three-way decomposition techniques, which aim at operating a dimension reduction, one of the most used is the CANDECOMP/PARAFAC decomposition. This approach assumes a decomposition as a sum of rank-1 terms. Herein, a rank-(Lr,Lr,1) decomposition is alternatively considered. Introduced by De Lathauwer in the context of signal processing, such an approach can handle more general situations. After a presentation of the rank-(Lr,Lr,1) decomposition, algorithmic aspects are detailed together with indices related to the final solution. Interestingly, connections between this decomposition technique and both STATIS and PARAFAC methods are highlighted. Finally, the method of analysis is illustrated on the basis of a dataset pertaining to sensory evaluation.

Keywords: Three-way decomposition, STATIS, PARAFAC, sensory.

Minimum RP estimators for independent but not identically distributed observations

Elena Castilla[1], María Jaenada[2], Leandro Pardo[3]

[1,2,3]Department of Statistics and OR, Complutense University of Madrid, Spain

In real life we often deal with independent but not identically distributed observations (i.n.i.d.o), for which the most well-known statistical model is the multiple regression model (MRM) without random covariates. While the classical methods are based on the maximum likelihood estimator (MLE), it is well known its lack of robustness to small deviations from the assumed conditions. In this paper, and based on the Rényi's pseudodistance (RP), we introduce a new family of estimators in case our information about the unknown parameter is given for i.n.i.d.o.. This family of estimators, let say minimum RP estimators (as they are obtained by minimizing the RP between the assumed distribution and the empirical distribution of the data), contains the MLE as a particular case and can be applied, among others, to the MRM without random covariates.

Keywords: Independent not identically distributed observations, Minimum Rényi's pseudodistance estimators, Robustness.

Reweighting with machine learning techniques in panel surveys. Application to the Health Care and Social Survey

Luis Castro-Martín[1], María del Mar Rueda[1], Carmen Sánchez-Cantalejo[2], Ramón Ferri-García, Andrés Cabrera-León[2]

[1]Department of Statistics and Operational Research. University of Granada, 18071, Spain, [2]Andalusian School of Public Health, Granada, Spain

Healthcare statistical services worldwide have used probability surveys to respond to such information needs. The Health Care and Social Survey (ESSOC) research project arises from the need to provide specific, reliable, early, and timely data on the impact of COVID-19 that can be considered when making decisions to prepare and provide an effective Public Health response in the different affected populations. The objective of this survey is to determine the magnitude, characteristics and evolution of the impact of COVID-19 on overall health and its socioeconomic, psychosocial, behavioral, occupational, environmental, and clinical determinants in the general population and that with greater socioeconomic vulnerability. This survey has an overlapping panel design. The use of panels address the problem of non-response. This problem is particularly aggravated in the case of panel surveys, due to the fatigue of the population when repeatedly surveyed. In this work, we test a new reweighting method that produces estimators that are suitable for survey data affected by non-response where auxiliary information exists on various levels. The weights are the result of a two-step process: the original sampling design weights are corrected during a 1st phase by modeling the non-response with respect to the longitudinal effective sample obtained in the previous occasion using state-of-the-art machine learning techniques. Then, during a 2nd phase, they are calibrated using the auxiliary information available at the population level. The proposed method is applied to the estimation of totals and proportions as well as the change in the ESSOC.

Keywords: reweighting, panel surveys, calibration, health surveys, machine learning methods.

Political Trust to National Institutions: The Significance of Items' Level of Measurement in the Validation of Constructs

Anastasia Charalampi[1], Eva Tsouparopoulou[2], Joanna Tsiganou[3], Catherine Michalopoulou[4]

[1]Postdoctoral Fellow, Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece, [2]Ph.D. Candidate, Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece, [3]Researcher A, The National Centre for Social Research, Athens, Greece, 4Professor of Statistics, Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece

The most important consideration in any statistical analysis is to ascertain the level of measurement of the input variables which guides the appropriate methodological steps to be used. In this paper, we carry out the validation of the 2008 European Social Survey (ESS) and European Values Study (EVS) measurement of political trust to national institutions for Greece, Portugal and Spain when items are considered as pseudo-interval (ESS) and ordinal (EVS). For the validation of this construct, the sample in each country was randomly split into two halves and first exploratory factor analysis (EFA) was performed on one half-sample in order to assess its construct validity. Secondly, the structure identified by EFA was investigated by carrying out confirmatory factor analysis (CFA) on the second half. Based on the full sample, the psychometric properties of the resulting scales were assessed. In all countries of both surveys, EFA performed on the first half-samples resulted in a unidimensional solution based on the four common items of the political trust to national institutions. CFA performed on the second half-samples and the full samples resulted in adequate model fit for all cases. Moreover, the analysis provided reliable scales and of adequate convergent validity. The methodology presented may be easily applied to other cases of validating scales comprised of pseudo-interval or ordinal items.

Keywords: Political trust, Exploratory factor analysis, Confirmatory factor analysis, Reliability, Validity.

The state of the art in flexible regression models for univariate bounded responses

Agnese Maria Di Brisco[1], Roberto Ascari[2], Sonia Migliorati2, Andrea Ongaro[2]

[1]Department of Studies in Economics and Business (DISEI), Università del Piemonte Orientale, Novara, Italy, [2]Department of Economics, Management and Statistics (DEMS), Università di Milano-Bicocca, Milano, Italy

Modeling bounded continuous responses, such as proportions and rates, is a relevant problem in methodological and applied statistics. A further issue is observed if the response takes values at the boundary of the support. Given that standard models are unsuitable, a successful and relatively recent branch of research favors modeling the response variable according to distributions that are well-defined on the restricted support. A popular and well-researched choice is the beta regression model and its augmented version. More flexible alternatives to the beta regression model, among which the flexible beta and the variance inflated beta, have been tailored for data with outlying observations, latent structures, and heavy tails. These models are based on special mixtures of beta distributions and their augmented version to handle the presence of values at the boundary of the support. The aim of this contribution is to provide a comprehensive review of the most popular existing models. Simulation studies and applications to real data are performed to show the relevance of correctly modeling the bounded response and to make comparisons among models. Inferential issues are dealt with by the (Bayesian) Hamiltonian Monte Carlo algorithm. Finally, the FlexReg package, a newly available tool on CRAN that offers efficient and easy-to-use implementation of regression models for bounded responses, is also described.

Keywords: FlexReg R package, mixture model, proportions, Bayesian inference.

Numerical Studies of Implied Volatility Expansions under the Gatheral Model

Marko Dimitrov[1], Mohammed Albuhayri1, Ying Ni[1], Anatoliy Malyarenko1
[1]Division of Mathematics and Physics, Mälardalen University, Västerås, Sweden

The Gatheral model is a three-factor model with mean-reverting stochastic volatility that reverts to a stochastic long-run mean. In this paper, the authors review and extend previous results on the first, second, and third-order implied volatility expansions under this model. Using Monte-Carlo simulation as the benchmark method, extensive numerical studies are conducted to investigate the accuracy and properties of these analytical expansions. Moreover, calibration of the model using these implied volatility expansions is conducted. The market data used in this calibration study include time series of market implied volatilities for an underlying market index and an underlying equity stock, for periods both prior to and during the COVID-19 crisis.
Keywords: Gatheral Model, Implied Volatility Expansions, Calibration.

Analytics of Portfolio Selection Dimensions: Return, Risk, Diversification

Yiannis Dimotikalis
Dept of Management Science and Technology (Agios Nikolaos), Hellenic Mediterranean University, Crete, Greece

The modern portfolio theory approach in portfolio selection, given the return of some financial assets, is risk minimization measured by the variance of associated portfolio. The associated return-variance efficient frontier represents the available solutions. The Sharpe Index (return/risk) is the appropriate portfolio selection criterion to balance the dimensions of return and risk. The known problem of this approach is the portfolio diversification, the concentration of optimum portfolio on few of the assets. The diversification which is a third dimension of portfolio can be measured by the entropy of portfolio asset weights. A return-entropy efficient frontier used where the maximization of portfolio weights entropy is the objective of optimization and naïve solution is the equal weights portfolio. In the

return-entropy efficient frontier Entropic Return index (return×entropy) is the proposed criterion to overcome the naïve equal weights solution. The assessment of those two portfolio efficient frontiers is the purpose of this paper. Using a dataset of 10 stocks the optimum portfolios, by the two efficient frontiers and related indexes, evaluated in the out of sample period performance. The observed performance differences, in short and medium term out of sample horizon, indicates that for the portfolio selection problem all the three dimensions are important.
Keywords: Maximum Entropy Frontier, Portfolio Selection, Equal Weights Portfolio, Portfolio Diversification.

A Possibilistic Approach for Subspace Clustering

Stavroula Eleftheraki[1], Konstantinos D. Koutroumbas[2], Athanasios A. Rontogiannis[2]
[1]Athena Research Center, Athens, Greece, [2]National Observatory of Athens, Athens Greece

Subspace clustering refers to the problem of clustering a collection of data points drawn from multiple subspaces of unknown dimension. A common problem when handling this type of data is the accurate determination of the clusters that lie around intersected subspaces and near to their intersection. In that vein, a novel possibilistic iterative algorithm was developed, called Adaptive Possibilistic K-subspaces (AP K-subspaces). Due to its possibilistic nature, where the “degree of compatibility” of a data point with a certain cluster is independent from its degree of compatibility with the other clusters, the algorithm is able to handle reliably clustering problems as mentioned above. A key feature of the proposed algorithm is that all of its parameters are adapted during its real-time execution. In addition, the algorithm does not require a priori knowledge of the exact number of clusters. In contrast, beginning with an overestimated number of clusters, AP K-subspaces has the ability (with suitable initialization) to gradually reduce it by eliminating nonphysical clusters and, finally, to end up with the correct number of clusters (and, of course the correct clustering of the data). Moreover, a low-rank approach, called Alternating Iteratively Reweighted Least Squares (AIRLS), can be adopted to the presented framework in order to estimate the subspaces where the

clusters live, in each iteration. Experiments on both synthetic and real data illustrate the effectiveness of the proposed algorithm.

Keywords: clustering-alternating minimization, cluster elimination, lowrank, subspace clustering, parameter adaptation, possibilistic clustering.

Covid-19 data imputation by Principal Component Prediction (PCP) models

Manuel Escabias[1], Christian Acal[1], Ana M. Aguilera[1], Mariano J. Valderrama[1]

[1]Department of Statistics and O.R., University of Granada, Granada, Spain

The first wave of the Covid-19 pandemic in Spain occurred between March and June 2020. In those early days of the pandemic, Spanish authorities published daily data of the evolution of the pandemic in the different territories in which Spain is organized (17 territories). The daily data of Number of confirmed cases, hospitalized people, admitted people in intensive care units (ICU), recovered people and deceased persons were published. All of these data were cumulatively published. A problem arose when some territories modified the way of communicating some data what made necessary for authorities to reconstruct the original data in order to make comparable all variables among all territories. The goal of this work is to use principal component prediction models (PCP models) to predict the cumulative curves of some of the variables (hospitalized people and people admitted to ICU) from the rest (cases, deceases and recoveries). A regularization of the data became necessary due to the absence of uniformity in the publication of the observations in the different territories, before model fits. The proposed principal component prediction model is a functional principal component based method for function on function linear regression analysis.

Keywords: Functional data analysis, Functional principal component analysis, function on function regression, PCP models.

Rate of Convergence of MLE of the drift in α-Brownian Bridge

Khalifa Es-Sebaiy[1], Jabrane Moustaaid[2]
Cadi Ayyad University, Marrakech, Morocco

We consider the alpha-Brownian bridge process X. An example of interesting problem related to X is the statistical estimation of drift parameter alpha when one observes the whole trajectory of X. A natural candidate is the maximum likelihood estimator (MLE). The aim of this communication is to provide, when alpha>1/2, an optimal rate of Kolmogorov distance for central limit theorem of the MLE.
Keywords: Alpha-Brownian bridge, Rate of convergence, MLE, Kolmogorov distance, Malliavin calculus.

A continuous time network evolution model governed by a branching process

Istvan Fazekas[1], Attila Barta[1]
[1]Faculty of Informatics, University of Debrecen, Debrecen, Hungary

A continuous time network evolution model is considered. The basic units of the model are triangles describing 3-interactions. The evolution of the network is governed by a continuous time branching process. The asymptotic behaviour of the model is studied. It is proved that the number of triangles, edges and vertices have the same exponential magnitude on the event of non-extinction. Besides mathematical proofs, simulation results support the results. Our theorems are extensions of the results [1].
References:
[1] Mori, T. F. and Rokob, S. A random graph model driven by timedependent branching dynamics. Annales Univ. Sci. Budapest., Sect. Comp. 46 (2017), 191-213.
Keywords: Network evolution, random graph, continuous time branching process, 3-interaction, Malthusian parameter, Poisson process, life time, extinction.

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

A Semi-Widely Linear Estimation Algorithm for Widely Factorizable Quaternion Signals

Rosa M. Fernández-Alcalá, Jesús Navarro-Moreno, Juan Carlos Ruiz Molina

[1]Department of Statistics and Operations Research, University of Jaén, Jaén, Spain

In the last decades, the use of hypercomplex signals has gained a great popularity in processing multidimensional signals. Among the different hypercomplex algebras, a special interest has been given to quaternions, due to their good properties of this algebra.

Quaternion signals can present different of improperness which requires different kinds of widely linear processing. In general, the optimal processing is the widely-linear processing, which requires to operate simultaneously on the quaternion signal and its three involutions. Notwithstanding, when the signal is Q-proper or $\mathbb{C}^{\eta}$–proper, the optimal processing is reduced to the strictly lineal (SL) or semi-widely linear (SWL) processing, respectively. Notice that SL processing ignores the involutions and SWL processing makes use of the quaternion and its involution over the pure unit quaternion η. In this paper, the estimation problem is addressed for a quarternion linear system with intermittent observations under $\mathbb{C}^{\eta}$–properness conditions. The signal to be estimated is assumed to be widely factorizable and the observation equation includes multiplicative noises given by a sequence of independent Bernoulli random variables. Based on a SWL processing, a linear mean squared recursive estimation algorithm is devised from correlation information. The proposed algorithm has its application in those situations where a state-space model is not readily at hand. Simulation examples illustrate the

better behavior of the SWL estimation algorithm presented here with respect to the conventional one obtained from a SL processing.

Keywords: $\mathbb{C}^{\eta}$–properness, Semi-widely Linear Processing, Quaternion Signals, Widely Factorizable Signals.

This work has been supported in part by I+D+i Project with reference number 1256911, under «Programa Operativo FEDER Andalucía 20014-2020», Junta de Andalucía, and the Project EI-FQM2 2019 of «Plan de Apoyo a la Investigación 2019-2020» of the University of Jaén.

Increase of retirement age and health state of population in Czechia

Tomáš Fiala[1], Jitka Langhamrová[2], Jana Vrabcová[3]

[1,2,3]Department of Demography, Faculty of Informatics and Statistics, Prague University of Economics and Business, Praha, Czechia

Population ageing is very often regarded to be a serious threat to the financial sustainability of welfare social systems, especially for the old-age pension system. Many states (including Czechia) intend to raise the retirement age above the standard usual threshold of 65 years. This threshold and its increase are usually determined by the economic or demographic point of view. There arises a natural question if there will be enough appropriate working positions for persons of older age and if the health state of these persons will enable them to work after reaching 65 years. The paper presents a brief analysis of the health state of the population in Czechia in the age since 60 until retirement age for various variants of increase of the retirement age in the future. Not only the healthy life expectancy but also e.g. lost years of healthy life and other indicators of such type are used for the analysis.

Keywords: Population ageing, retirement age, health state, healthy life expectancy, lost years of healthy life.

Performance Persistence of Polish Investment Funds: Mobility Measures

Dariusz Filip
Faculty of Social and Economic Sciences, Department of FinanceCardinal Stefan Wyszynski University in Warsaw (UKSW)

The purpose of this paper is to evaluate the phenomenon of performance persistence in a developing market. The analysis is conducted for Polish investment funds from three-time perspectives: monthly, quarterly, and yearly. The research approach applied is a Markovian framework supported with a few mobility measures. The results reveal the existence of limited performance persistence, which decreases as the time frame increases. Moreover, the observed propensity for a relative repetition of mutual fund performance in consecutive periods seems to involve losers rather than winners, and hence it takes the form of the "icy hands" effect.
Keywords: Markov chain, persistence, performance inertia, developing markets.

Blackjack and the Kelly bet: a simulation assessment of selected playing strategies

Jim Freeman[1], Haoyu Miao[2]
[1,2]Alliance Manchester Business School, University of Manchester, Manchester, UK

The paper is concerned with the application of the Kelly criterion to the casino game, blackjack. The comparative performance of a number of playing strategies used in this game is assessed using Monte Carlo simulation against a range of alternative bet size decisions.
Playing strategies considered include "Blind gambler, "Never bust", "Imitating the dealer" (Keren and Wagenaar, 1985).
The review yields many interesting insights – not all necessarily in favour of Kelly.
Keywords: Doubling up, Half Kelly bet, Kelly bet, Minimum bet.

Invariant description for batch version of UCB strategy with unknown control horizon

Sergey Garbar
Yaroslav-the-Wise Novgorod State University, ul. Bolshaya Sankt-Peterburgskaya, Russian Federation

We consider a batch processing variation of UCB strategy for multi-armed bandits with a priori unknown control horizon size n. Random rewards of the considered multi-armed bandits can have a wide range of distributions with finite variances. We consider batch processing scenario, when the arm of the bandit can be switched only after it was used a fixed number of times, and parallel processing is also possible in this case. A case of close distributions, when expected rewards differ by the magnitude of order N–1/2 for some fairly large N, is considered as it yields the highest normalized regret. Invariant descriptions are obtained for upper bounds in the strategy and for minimax regret. We perform a series of Monte-Carlo simulations to find the estimations of the minimax regret for multi-armed bandits. Maximum for regret is reached for n proportional to N, as it expected based of obtained descriptions.
The reported study was funded by RFBR, project number 20-01-00062.
Keywords: multi-armed bandit, UCB, batch processing.

Non-decreasing Subsequences for Exploring Statistical Independence

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

We show how the longest non-decreasing subsequence, identified in the graph of the paired marginal ranks of the observations, allows the construction of a statistic to develop an independence test in bivariate random vectors, see [3]. The random variables requirements do not include the continuity or the discrete condition, so this procedure expands

the proposal introduced in [1] and in [2], working either in the continuous and discrete case. We show the efficiency of the procedure in detecting dependence in real cases and through simulations.

References:

[1] García J.E.; González-López, V.A. Independence tests for continuous random variables based on the longest increasing subsequence. Journal of Multivariate Analysis, 2014, 127, 126-146 (Doi:10.1016/j.jmva.2014.02.010)

[2] García, J.E.; González-López, V.A. Independence test for sparse data. In AIP Conference Proceedings, 2016; Volume 1738, No. 1, p. 140002 (Doi: 10.1063/1.4951925)

[3] García, J.E.; González-López, V.A. Random Permutations, Non-Decreasing Subsequences and Statistical Independence. Symmetry 2020, 12, 1415 (Doi: 10.3390/sym12091415)

Keywords: Permutations; Hypothesis tests.

Language Modality Representation through Multiple Partition Markov Models

J.E. García, V.A. González-López, G.H. Tasca, V.G. Santos, F.R. Fernandes-Svartman

University of Campinas, Campinas, São Paulo, Brazil

In this paper, we propose a method to build a stochastic profile of a language's modality. Through that profile, we compare two modalities perceived as different by human hearing, M1 and M2, in order to decide if those modalities can be characterized as being really different by their stochastic performance, see [1]. This can be made by means of the Fundamental Frequency (F0), which can be linked with a discrete and finite alphabet A, containing the coding of events of interest (variations on F0). Records obtained from speakers reproducing each modality compound the set of independent samples, considered as coming from a set of processes, collection 1 (C1), related to modality M1, and collection 2 (C2), related to modality M2. Given a collection {(X {jt}), j=1,2,...,p} of independent discrete-time Markov processes with finite alphabet A and state space S, we state that the elements (i, s) and (j, r) in {1, 2,...,! p} × S are equivalent if and only if they share the same transition probability for all the elements in the alphabet. The equivalence allows reducing the

number of parameters to be estimated in the model avoiding deleting states of S to achieve that reduction, see [2]. This model allows us to represent the underlying and common stochastic law of the set of sequences. Through the equivalence relationship, we build global profiles for collection C1 and C2, respectively. The equivalence classes define an optimal partition of {1, 2,...,p} × S, and it is in relation to this partition that we define the profile of each modality.

References:

[1] García, J. E., González-López, V. A., & Tasca, G. H. (2020). Partition Markov Model for Covid-19 Virus. 4open, 3, 13.

[2] Cordeiro, M. T. A., García, J. E., González-López, V. A., & Mercado Londoño, S. L. (2020). Partition Markov model for multiple processes. Mathematical Methods in the Applied Sciences, 43 (13), 7677-7691.

Keywords: Multiple Partition Markov Models; Bayesian Information Criterion; Transition probabilitiesal.

The roll of the practice of statistics and statistical reasoning in health sciences students

Jaime Gaviria, Difariney González-Gómez, Jhony Villa-Ochoa

University of Antioquia, Faculty of Education, Medellín, Colombia

This paper summarizes part of the theoretical framework for situated statistical reasoning in postgraduate's health sciences from a university in south America. The aim of this work is to show how the practice of statistics in health sciences contexts is an important and necessary component of statistical reasoning. The framework for situated statistical reasoning is founded in three interrelated components: statistical reasoning, situated reasoning and the practice of statistics in health contexts. In this paper only the last component is described. The practice of statistics is founded in general and specific skills that health sciences students usually need in problem solving contexts. The different rolls that students play in their relationship with the practice of statistics are also discussed. As results was found that health sciences students have a different relation with the practice of statistics: for one side there are

professional users of statistics, which include health workers. This kind of users need to understand medical research results, apply statistical methods in their own field of work or research and often translate statistical results to other users. On the other hand, literature identifies specialist's users of statistics. These kinds of users are producers of statistical studies and generally are biostatistics or epidemiologists.

Keywords: Statistical reasoning, practice of statistics, health sciences, statistics education research.

Data Envelopment Analysis and Markov Models

Andreas C. Georgiou[1], Emmanuel Thanassoulis[2], Alexandra Papadopoulou[3]

[1]Quantitative Methods & Decisions Analytics Lab, Department of Business Administration, University of Macedonia, Thessaloniki, Greece, [2]Aston University, Aston Business School, Birmingham, UK [3]Aristotle University of Thessaloniki, Department of Mathematics, Thessaloniki, Greece

This paper analyses a series of hybrid models that embed Data Envelopment Analysis, within Markov models. As a vehicle for this experiment we use the attainability problem and we measure the efficiency of candidate recruitment policies to drive a system towards a desired target. It builds upon existing models in manpower planning and brings into the analysis a novel approach based on a non-parametric efficiency evaluation method. The alternative recruitment policies and/or alternative initial population structures are treated as Decision Making Units under uncertainty and are compared against an efficient frontier formed through a production possibility set defined by convex combinations of alternative policies. Various forms of the hybrid model are proposed and a numerical illustration is used to support the applicability of the model. The paper concludes with ideas of further research including both directions of the analysis that is DEA and markov modelling.

Keywords: Markov models, Data Envelopment Analysis, efficiency attainability.

A new non-monotonic link function for Beta regressions

Gloria Gheno
Ronin Institute, Montclair, NJ, USA

Beta regression is used to analyse data whose value is within the range (0,1), such as rates, proportions or percentages, and therefore is useful for analysing the variables that affect them ([1], [2]). This method is based on the beta distribution or its reparametrizations, proposed by [1] and by [3], to obtain a regression structure on the mean that is easier to analyse and interpret. For the regression for binary data, the literature has debated the problem of incorrect link functions and therefore proposed new links, such as gev (Generalized extreme value), while for the mean of the beta regression, the traditional link functions for binary regressions were used, i.e. logit, probit and complementary log-log. In this paper I propose a new link function for the mean parameter of a beta regression which has as its particular cases logit, representing a traditional symmetric link function, and gev, proposed for binary data due to its asymmetry. The new link function, proposed by me, has the advantage that it can also be non-monotonic, unlike those proposed until now. The parameters are estimated maximizing the likelihood function, using a modified version of the genetic algorithm, giving so greater importance to traditional link functions than the others. I compare my method with the one proposed by [1], in which the researcher decides a priori the link function, using simulated data, so as to be able to compare which of the two methods is closest to the true values. My method, therefore, is better because it is able to correctly determine the link function with which the data was simulated and to estimate the parameters with less error.

References:

[1] FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. Journal of applied statistics, 2004, 31.7: 799-815.

[2] SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. Computational Statistics & Data Analysis, 2010, 54.2: 348-366.

[3] CRIBARI-NETO, F.; SOUZA, T. C. Testing inference in variable dispersion beta regressions. Journal of Statistical Computation and Simulation, 2012, 82.12: 1827-1843.

Keywords: beta regression, genetic algorithm, likelihood, link function.

Response clustering in component-based GLM

Julien Gibaud, Xavier Bry, Catherine Trotier
IMAG, Montpellier, France

Originally, the Supervised Component-based Generalized Linear Regression (SCGLR) methodology, proposed by Bry et al. (2013), was designed to find explanatory components in a large set of possibly highly redundant covariates. This methodology optimizes a trade-off criterion between the model's Goodness-of-Fit (GoF) and some Structural Relevance (SR) (Bry et al. (2015)) of directions with respect to the explanatory variables. This methodology allows both to find strong explanatory directions and to produce regularized predictors compatible with the high-dimensional framework. However, SCGLR assumes that all the responses are explained by the same explanatory dimensions. To overcome this limitation, we propose to extend this methodology to mixture models of the outcomes, enabling it to identify clusters of responses dependent on common explanatory components. Our work is based on the modeling approach of Dunstan et al. (2013) who propose using Finite Mixture Models (FMM) to analyze data communities, which they call Species Archetype Model (SAM). The idea is to assume that the responses (species) can be clustered into a small number of groups. In the spirit of Gibaud et al. (2020), we propose to cluster outcomes in groups predicted by the same supervised components built by SCGLR. In an ecological framework for instance, communities of species should be modeled by components characteristic of each community. The flexibility of our method allows to mix several probability distribution functions from the exponential families for the outcomes, dealing thus with presence-absence, count or biomass data. In order to estimate the model parameters, we implement an algorithm which alternatively evaluates the mixture parameters with the Expectation-Maximization (EM) algorithm and then finds the supervised components with the Projected Iterated Normed Gradient (PING) algorithm. For the sake of clarity, a few simple simulation studies illustrating the interest of our model will be presented.
Keywords: Supervised components, response mixture, EM algorithm, clustering.

Goodness--of--fit procedures for compound distributions with an application to Insurance

P.-O. Goffard, S. Rao Jammalamadaka, S. G. Meintanis[1]
[1]*National and Kapodistrian University of Athens Athens, Greece*

Goodness-of-fit procedures are introduced for testing the validity of compound models. New tests that utilize the Laplace transform as well as classical tests based on the distribution function are investigated. A major area of application of compound laws is in insurance, to model total claims resulting from specific claim frequencies and individual claim sizes. Monte Carlo simulations are used to compare the different test procedures under a variety of specifications for these two components of total claims. Also a detailed application to an insurance data set is presented.

A Generalized Mean under Non-regular Frameworks and Extreme Value Index Estimation

M. Ivette Gomes[1], **Lígia Henriques Rodrigues**[2], **Dinis Pestana**[3]
[1,3]*Centro de Estatística e Aplicações (CEAUL), Universidade de Lisboa, Lisboa, Portugal,* [2]*Centro de Investigação em Matemática e Aplicações (CIMA), Universidade de Évora, Évora, Portugal*

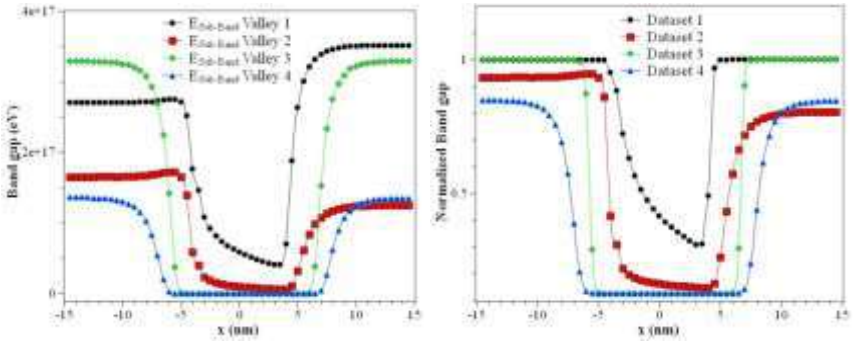
The Hill estimator, one of the most popular *extreme value index* (EVI) estimators under a heavy right-tail framework, i.e. for a positive EVI, denoted by ξ, is an average of the log-excesses, $V_{ik} := \log(X_{n-i+1:n}) - \log(X_{n-k:n})$, $1 \le k < n$-1, with $X_{i:n}$, $1 \le i \le n$, denoting the i-th ascending order statistic associated with an available sample, $(X_1, \ldots, X_n)$. Consequently, the Hill EVI-estimator can be regarded as the logarithm of the geometric mean, i.e. the mean of order $p = 0$ of $U_{ik} := \exp(V_{ik}) = X_{n-i+1:n} / X_{n-k:n,}$, $1 \le k < n$-1. We can thus more generally consider any real p, the *mean of order p* (MO_p) of those same statistics and the associated Mo_p EVI-estimators, sometimes called harmonic moment EVI-estimators. The normal asymptotic behavior of these Mo_p EVI-estimators has been obtained for p $p < 1/(2\xi)$, with consistency achieved for $p < 1/\xi$. The non-regular framework, i.e. the case $p \ge 1/(2\xi)$, will be now considered. For $p = 1/(2\xi)$,

asymptotic normality is still achieved, but with a slower rate of convergence. If $1/(2\ \xi) < p \leq 1/\xi$, i.e. $1 < \alpha = 1/(p\ \xi) \leq 2$, a sum-stable limit, with characteristic exponent α,, appears. Consistency is no longer achieved for $p > 1/\ \xi$, but an almost degenerate behavior appears for $p > 1/\ \xi$, An algorithm providing an almost degenerate Mo_p EVI-estimation is suggested and studied for finite samples through Monte-Carlo simulations. Applications to simulated and real data are further performed.
Keywords: Generalized means; non-regular frameworks; statistics of extremes.

A method of Big Data Collection and Normalization for Electronic Engineering Applications

Naveenbalaji Gowthaman[1], Viranjay M. Srivastava[2]
[1,2]Department of Electronic Engineering, Howard College, University of KwaZulu-Natal, Durban, South Africa

The data collection and storage has become the greatest challenges and tedious processes in the data science engineering. The data from various nodes (sensors, bridges, switches, hubs, etc.,) in the environment or a particular system is collected at the nodes from which they get to the storage point. These types of operations needs separate workforce to monitor the whole process of data handling. This proposed work mainly focuses on the data analytics of creating a normalized data from the unprocessed data. This reduces the manipulation of data when it is of different form. The data may be realistic depending on the system which produces it. The normal distribution applies to the collected data to create a dataset which is distributed over the continuous probability density function. It extends up to infinity in both the direction of axes. The proposed work provides easy storage and data retrieval method in case of large data volumes. The proposed data recovery is compliant than the conventional data collection methods. This type of data interpretation provides security and confidentiality of the user's data.



(a) Data derived from the simulation of MOSFET (b) Normalized data for interpretation
Fig. Data Collection and its Normalization for better analysis

Keywords: Electrical field, Data Storage, Data Interpretation, segmentation, normalized data, standard deviation.

Stochastic Runge–Kutta solvers based on Markov jump processes and applications to nonautonomous systems of differential equations

Flavius Guias
Dortmund University of Applied Sciences and Arts, Sonnenstr. 96, 44139, Dortmund, Germany

In this presentation we will discuss the stochastic Picard–Runge–Kutta solvers for autonomous systems of ordinary differential equations introduced in [1], [2] and [3] and introduce a modification of them, which makes them suitable for nonautonomous systems.
The principle used by this family of schemes is based on the connection between the infinitesimal generators of Markov jump processes and corresponding differential equations. The step function $X^{\sim}$ computed by simulating the jump processes can serve as a predictor which is further improved by suitable correction steps. Given the improved approximation $X^{*}(t)$ at time t, we compute the corresponding approximation at time $t + h$ by an integral scheme of the form

$$X^*(t+h) = X^*(t) + \int_t^{t+h} Q(s)\,ds.$$

For computing the improved approximations $X*(\cdot)$ we take for the integrand Q a polynomial which interpolates some equidistant intermediate values of $F(X^{\sim}(\cdot))$ between t and $t + h$. According to [1], [2] and [3], by using an exact quadrature formula in order to compute the integral above, we can employ in this stochastic context the same principle of the deterministic Runge–Kutta method in order to compute a better approximation for the solution of the equation. The final result is a high precision scheme with several layers, which starts from the crude approximation delivered by the standard jump process, and based on this data it computes several steps in which the approximations are successively refined.

References:

1. F. Guia¸s and P. Eremeev. Improving the stochastic direct simulation method withapplications to evolution partial differential equations. Appl. Math. Comput., 289 : 353–370, 2016.

2. F. Guia¸s. Stochastic Picard–Runge–Kutta Solvers for Large Systems of Autonomous Ordinary Differential Equations. 2017 Fourth International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), DOI 10.1109 / MCSI.2017.55 : 298–302, 2017.

3. F. Guia¸s. High precision stochastic solvers for large autonomous systems of differential equations.Int. J. Mathem. Models. Meth. in Appl. Sci., 13 : 60–63, 2019.

Performance of Evaluation of Diagnosis of Various Thyroid Diseases Using Machine Learning Techniques

Burcu Bektas Gunes[1], Evren Bursuk[2], Ruya Samli[3]

[1]Department of Computer Engineering, National Defense University, Turkish Naval Academy, Tuzla, Istanbul, Turkey, [2]Department of Biomedical Device Technology, Istanbul University-Cerrahpasa, Beylikduzu, Istanbul, Turkey, [3]Department of Computer Engineering, Istanbul University-Cerrahpasa, Avcılar, Istanbul, Turkey

Thyroid cancer is the second cancer type which is prevalent among women in Turkey. The number of people diagnosed with thyroid cancer is estimated as 44,280 in the United States in 2021 according to the report

published by the American Cancer Society. Early diagnosis and treatment are very important for preventing this disease. This study focused on predicting 5 different thyroid diseases based on various symptoms and reports of the thyroid. Several machine learning algorithms such as Support Vector Machine, k-Nearest Neighbors, Linear Regression, Naive Bayes, Decision Trees, Artificial Neural Network are used for diagnosis of various thyroid diseases and their classification performances are compared with each other. For this purpose, thyroid disease dataset gathered from the Istanbul University-Cerrahpasa Nuclear Medicine Department and Endocrinology Departments was used.
Keywords: Thyroid Disease, Machine Learning Algorithms, Prediction Model, Medical Diagnosis, Artificial Neural Networks.

Interpreting a Topological Measure of Complexity for Decision Boundaries

Alan Hylton[1], Ian Lim[2], Michael Moy[3], Robert Short[4]
[1,4]NASA Glenn Research Center, Cleveland Ohio, United States, [2]The University of Texas at Arlington, Arlington Texas, United States, [3]Colorado State University, Fort Collins, Colorado, United States

We propose a method to examine the decision boundaries of classification algorithms to yield insight into the nature of overfitting. In machine learning, model evaluation can be performed via two common techniques: train-test split or cross validation. In this paper, we grow this toolkit to include tools from the field of topological data analysis. In particular, we use persistent homology, which roughly characterizes the shape of a data set. Our method focuses on binary classification, using training data to sample points on the decision boundary of the feature space. We then calculate the persistent homology of this sample and compute metrics to quantify the complexity of the decision boundary. Our experiments with data sets in various dimensions suggest that in certain cases, our measures of complexity are correlated with a model's ability to generalize to unseen data. We hope that refining this method will lead to a better understanding of overfitting and a means to compare models.
Keywords: Topological Data Analysis, Machine Learning, Classification Algorithm.

Coupon collector problem with universal coupon

Jelena Jocković[1], Bojana Todić[2]
[1,2]*Faculty of Mathematics, University of Belgrade, Belgrade, Serbia*

The coupon collector problem (CCP) is a mathematical model that belongs to the family of urn problems, and can be formulated as follows: a company issues coupons of different types (say, elements of the set $N_n = \{1,2, \ldots, n\}$), each with a particular probability of being issued. The object of interest is the number of coupons that must be collected to obtain a full collection. We consider the following generalization of CCP: We assume that, apart from elements of N_n, the set of coupons consists of a null coupon (which does not belong to the collection) and, additionally, a universal coupon (joker), an element that can replace any of the elements from the set N_n, one at a time. We assume that sampling is with replacement, coupon $k \in N_n$ is drawn with probability p_k, joker is drawn with probability p_J, null coupon is drawn with probability p_N, and p_N, $p_J < 1$. Let $W_{n,c}$ be the waiting time until $m \leq c$ elements of N_n are sampled, and $c - m$ jokers are sampled, $1 \leq c \leq n$. In other words, $W_{n,c}$ is the waiting time until a given portion of the collection N_n is sampled, although some, or all of the elements can be replaced by jokers. In this paper, we obtain the distribution of $W_{n,c}$, and provide some immediate consequences of that result. We also determine the distribution that stochastically minimizes the waiting time $W_{n,c}$, when p_J is fixed. Coupon collector problem and its variants have well known applications in the field of engineering, and, recently, in biology.

Keywords: Coupon collector problem, Waiting time, Universal coupon.

Exploring Excess Mortality during the Corona Pandemic with Seasonal ARIMA Models

Karl-Heinz Jöckel[1], Peter Pflaumer[2]

[1]Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Germany, [2]Fakultät Statistik, Technische Universität Dortmund, Germany

Excess deaths will be defined as the difference between the observed numbers of deaths in a specific period and expected numbers of deaths in the same period. Excess death estimates can be calculated in several ways, and will be different depending on the methodology and assumptions about how many deaths are likely to occur. Often, a simple and easy to understand approach is chosen where excess deaths are calculated by comparing the current deaths with the average across previous years. A more complex and sophisticated approach is model based (cf., e.g., the models of Euromomo or of the Centers for Disease Control and Prevention – CDI). We suggest an alternative model. A seasonal autoregressive moving average model (SARIMA) is applied. Model identification and selection is based on autocorrelation and partial autocorrelation functions. After parameter estimation model checking is carried out by testing whether the estimated model conforms to the specifications. Excess deaths for selected European countries and the United States are calculated as the difference between the observed count and one of two thresholds- either the expected value or the upper 95% prediction interval. It turns out that these countries were affected by excess mortality very differently in 2020. The number of deaths each year depends both on death probabilities and on the size and the age structure of a population. Our proposed model implicitly takes these changes into account.

Demography and policies in V4 countries

Michaela Kadlecová[1], Filip Hon[2], Jitka Langhamrová[3]
[1,2,3]Department of Demography, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic

The article deals with the analysis of the recent and currently planned policies with their meaningfulness based on taking into account demographic development in the countries of the Visegrad Group. The Visegrad Group (V4) is an alliance of four Central European countries, including the Czech Republic, Hungary, Poland and Slovakia. In addition to analyzing the current and recent situation, the paper also deals with estimates of future fertility and mortality levels in these countries. It shows there are quite significant differences in the political laws in the natalist policy and in the pension system of individual countries and, at the same time, places them in the context of the future further aging of populations which can be, based on projections, expected in all of the Visegrad countries.
Keywords: Natalist policy, Pension reform, Projection, Second demographic transition.

The use of Penalized Regime Switching for the Modeling of Consultation Rate Data

Emmanouil-Nektarios Kalligeris[1], Alex Karagrigoriou[1], Christina Parpoula[2]
[1]Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, Samos, Greece
[2]Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece

Regime switching in conjunction with penalized likelihood techniques could be a robust tool concerning the modeling of dynamic behaviors of consultation rate data. To that end, in this work we propose a methodology that combines the aforementioned techniques, and it is tested through a real case and a simulation study.

References:
Kalligeris EN, Karagrigoriou A and Parpoula C. On Mixed PARMA Modeling of Epidemiological Time Series Data. Commun Stat Case Stud Data Anal 2019; 6(1): 36-49.
Kalligeris EN, Karagrigoriou A and Parpoula C. Periodic-Type Auto-Regressive Moving Average Modeling with Covariates for Time-Series Incidence Data via Changepoint Detection. Stat Meth Med Res 2019; https://doi.org/10.1177/0962280219871587.
Keywords: Cross-validation, Elastic net, Markov switching, Penalized likelihood, Variable selection.

Stochastic Orders and Bounds for Ruin Theory

Lazaros Kanellopoulos
PhD student, Department of Statistics and Insurance Science, University of Piraeus, Greece

We derive new results concerning stochastic comparisons of the distribution of the deficit in the renewal risk model and the function $K_x(y)$ introduced by Willmot and Cai (2004). Finally, some new results about the NBUL class of life distributions are obtained.
References:
Willmot, G. E., Cai, J.(2004). On application of residual lifetimes of compound geometric distributions. J. Appl. Prob. 41, 802-815.
Willmot, G. E. (2002) Compound geometric residual lifetime distributions and the deficit at ruin. Insurance: Mathematics and Economics 30, 421-438.
Psarrakos, G. (2010) A note on convolutions of compound geometric distributions. Statistics and Probability Letters 79, 1231-1237.
Belzunce, F., Ortega, E. M., Ruiz, J. (2001). A note on stochastic comparisons of excess lifetimes of renewal processes. J. Appl. Prob. 38, 747-753.
Keywords: Ruin Theory, Renewal processes, excess lifetime, NBUL, ageing classes.

Machine Learning's Dropout Training is Distributionally Robust Optimal

Yang Kang
Ph.D in statistics from Columbia University

This paper shows that dropout training in Generalized Linear Models is the minimax solution of a two-player, zero-sum game where an adversarial nature corrupts a statistician's covariates using a multiplicative nonparametric errors-in-variables model. In this game---known as a Distributionally Robust Optimization problem---nature's least favorable distribution is dropout noise, where nature independently deletes entries of the covariate vector with some fixed probability delta. Our decision-theoretic analysis shows that dropout training---the statistician's minimax strategy in the game---indeed provides out-of-sample expected loss guarantees for distributions that arise from multiplicative perturbations of in-sample data. This paper also provides a novel, parallelizable, Unbiased Multi-Level Monte Carlo algorithm to speed-up the implementation of dropout training. Our algorithm has a much smaller computational cost compared to the naive implementation of dropout, provided the number of data points is much smaller than the dimension of the covariate vector.
This is joint work with Jose Blanchet, Jose Luis Montiel Olea, Viet Anh Nguyen, and Xuhui Zhang.

Software Cost Estimation Using Machine Learning Algorithms

Sukran Ebren Kara[1,2], Ruya Samli[2]
[1]Department of Computer Technologies, Cizre Vocational School, Sirnak Univesity, Sirnak, Turkey, [2]Department of Computer Engineering, Istanbul University–Cerrahpasa, Avcilar, Istanbul, Turkey

Software cost estimation is one of the most important problems of software projects. The project manager reduces the ambuguities in the project by estimating the project cost correctly. Otherwise, serious economic problems will arise. As a result of the growth and complexity of software projects, new cost estimation methods are constantly being developed. In this study, the cost of software projects is tried to be estimated by using machine learning algorithms. Project cost estimation has been made by

testing in different machine learning algorithms with Wakiato Environment for Knowledge Analysis (WEKA) data mining software tool. Algorithms were applied to China dataset taken from PROMISE data store with 10-fold cross validation technique and results, performance criterion correlation coefficient, error rates mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE). Thanks to this study, the information about which algorithms can be used for software cost estimation, what the estimation results might be when these algorithms are applied to China dataset and which algorithm is the best working have been reached.
Keywords: Machine Learning Algorithms, Software Cost Estimation, WEKA, China Dataset.

An Exponentiality GoF Test vs. Light & Heavy Tailed Alternatives

Alex Karagrigoriou[1], Ilia Vonta[2]
[1]Laboratory of Statistics and Data Analysis, University of the Aegean,
[2]Department of Mathematics, National Technical University of Athens, Greece

Log-concavity is associated with light-tail distributions while logconvexity with heavy-tailed ones (see An 1998). Log-concavity is a widely studied topic with substantial literature (see e.g. Saumard 2014; Gupta and Balakrishnan 2012) which is mainly due to the fact that logconcavity provides desirable estimation properties (see Walther 2009; Dumbgen and Rufibach 2009). This work is filling up the gap in the literature regarding the verification of the log-concavity property. At the same time it is vital in reliability, in engineering and in stochastic modeling for distinguishing between an exponential distribution, a lighttailed one or a heavy tail one. In this work we propose a goodness of fit exponentiality test to be used for distinguishing between exponential and light or long-tail distributions. The test statistic is based on the principle of the so called, single big jump. The details of the formulation of the test are provided, an extended simulated study which shows the performance of the proposed test statistic is given, and some concluding remarks are stated.
Keywords: Exponentiality test, Goodness of fit test, Log-concavity, Log-convexity, Single big jump principle.

Mathematical models for the prediction of Covid-19 cumulative confirmed incidents in Greece

Giannis A. Kechagias
School of Economics, Faculty of Economics and Political Sciences, Aristotle University of Thessaloniki, Greece

The novel coronavirus pandemic has become a major threat to the whole world. The health policy makers need estimates of the confirmed cases in future to make informed decisions about the required interventions. Growth models are used to describe a particular type of growth of biological systems over time, and they are appropriate in modelling the growth of an epidemic. The ARIMA models are widely used statistical approaches for time-series analysis and forecasting. In this work we present two of the most commonly used growth models, namely, Logistic model and Gompertz model. We also present two variations of those models the Log-logistic model and the Sloboda model. The objective is to predict the cumulative cases of covid-19 in Greece starting from observed data, using a forecast based on the above-mentioned growth models A Non-Linear Least Squares method was employed to estimate the parameters of each growth model such that the model curve fits best the given data in the least squares sense. Estimation was completed using the MATLAB package. This work also aims to analyze the time-series of cumulative cases in Greece, using the Box-Jenkins methodology, and find the best ARIMA model for predicting the future number of cumulative cases. The building of ARIMA model is handled by statistical package EViews11. A comparison between the models is also provided.
Keywords: Logistic model, Gompertz model, Non-Linear Least Squares method, ARIMA model, pandemic.

Multivariate Bernoulli-Geometric Stable Distributions

Yury Khokhlov[1], Victor Korolev[2]

[1,2]Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Department of Mathematical Statistics, GSP-, 1-52, Leninskie Gory, Moscow, Russia

Random sums are very popular in many applied problems. Very often it is assumed that the summation index has a geometric distribution and the summands are independent and identically distributed. We consider a new model in which the summands have distributions belonging to some finite set and are chosen randomly according to some multivariate Bernoulli distribution (see [1]). The limit distributions in such a model have some special form, which we called (by analogy with geometric stable distributions (see [2])) Bernoulli-geometric stable distributions. We introduce a more general definition of the multivariate Linnik distribution and find a connection between this distribution and the multivariate stable, Mittag-Leffler and Weibull distributions. We prove some characterization problem and consider several applications this model in actuarial mathematics and teletraffic theory. This research was carried out in accordance with the scientific program of the Moscow Center for Fundamental and Applied Mathematics and Faculty of Computational Mathematics and Cybernetics in Moscow University.

References:

[1] Yu. S. Khokhlov. Multivariate analogs of classical univariate discrete distributions and their properties. – Journal of Mathematical Sciences, Vol. 234, No. 6, November, 2018, 802-809.

[2] Kozubowski, T. J.; Rachev, S. T. Multivariate geometric stable laws. Journal of Computational Analysis and Applications, 1999, 1(4) 349-385.

Keywords: Multivariate random sums, Multivariate geometric distribution, Scale mixtures, Subordinated processes.

Concentration of the strong chromatic number of a random hypergraph

Alina Khuzieva[1], Tatiana Matveeva[2], Dmitry Shabanov[3]
[1]HSE University, Faculty of Computer Science, Moscow, Russia, [2]Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Moscow, Russia, [3]Moscow Institute of Physics and Technology, Laboratory of Combinatorial and Geometric Structures, Moscow Region

The work deals with estimating the threshold for the strong r-colorability of a random k-uniform hypergraph in the binomial model H(n,k,p). Recall that H(n,k,p) is revealed as Bernoulli scheme on k-subsets of an n-element vertex set. A vertex coloring of a hypergraph is said to be strong if any two vertices of the same color do not share a common edge. It is known that the threshold corresponds to the sparse case, when the expected number of edges is a linear function of n, cn, and c>0 depends on r and k, but not on n. The main result of the paper provides tight bounds for threshold as estimates for the parameter c. We establish the threshold up to a function that tends to zero with growth of k. The obtained results show that in the sparse case the strong chromatic number of H(n,k,p) is concentrated in one or two values with probability tending to 1.The proof is based on the second moment method, for its application, we also solve some extremal problem for doubly stochastic matrices. The research of the first and the third authors was funded by RFBR project number 20-31-70039. The research of the third author was also supported by grant of the President of Russian Federation no. MD-1562.2020.1.
Keywords: Random hypergraphs, Strong colorings, Second moment method, Doubly stochastic matrices.

3D Electron Beam Current Distribution Modelling

Elena Koleva[1,2], Lilyana Koleva[1]
[1]Institute of Electronics, Bulgarian Academy of Sciences, 1784 Sofia, 72 Tzarigradsko shoes blvd., Bulgaria, [2]University of Chemical Technology and Metallurgy, 1756 Sofia, 8 Kl. Ohridski blvd., Bulgaria

The electron beam (EB) technological processes like EB welding, EB additive technologies, etc. depend strongly on the characteristics of the electron beam, generated by the electron gun of each EB installation. The

3D EB radial and angular current distributions, defined for each transverse cross-section of the beam and different process parameters, determine the beam profile and are used to determine the invariant characteristics of the beam quality - brightness and emittance. There are several approaches for EB characterization, which generally can be divided into two big groups: i) assuming Gaussian distribution of the beam current density and ii) without this assumption. In this paper a tomographic technique for reconstruction of a two-dimensional object image from a set of its one-dimensional projections at different scanning angles is implemented for the estimation of 3D radial EB current distribution based on experimental measurements. There the Fourier transformation from real to frequency space and the consequent two-dimensional inverse Fourier transformation permit to reconstruct the beam cross-section current density image with their asymmetry features. Different modelling approaches are compared for the estimation of 3D radial EB current distribution (neural, kriging, spline). The estimated 3D models are used for the evaluation of the angular EB current standard deviations, beam profile (envelope), focus position and emittance. The influence of two process parameters - the focusing current and the EB current is investigated. The characterization of the electron beam is one of the necessary conditions for the transfer of technologies from one equipment to another, as well as for the comparison of the quality of different electron beam facilities (guns).
Keywords: Electron Beam, Radial and Current Density Distributions, Tomographic Approach.
Acknowledgements: The research was conducted within the framework of the project KP-06-N27/18, funded by the National Fund “Scientific Investigations”.

Quality Improvement of Quantitative and Bivariate Characteristics by Robust Optimization

Elena Koleva[1,2], Lilyana Koleva[1], Dmitriy Trushnikov[3]
[1]Institute of Electronics, Bulgarian Academy of Sciences, Sofia, Bulgaria,
[2]University of Chemical Technology and Metallurgy, Sofia, Bulgaria,
[3]Perm National Research Polytechnic University, Komsomolskiy Prospect, Russian Federation

The model-based robust approach is successfully applied to different industrial processes for improving their quality performance characteristics. The Robust engineering approach is implemented in cases of heteroscedasticity of the performance characteristics in presence of errors in the process parameters or noise factors. For its implementation, two models for each of the output characteristics are estimated - for their mean values and their variances. In this paper an electron beam welding case-study, which considers both quantitative and bivariate output characteristics, is presented. For the modelling of the dependence of the bivariate quality characteristic on the process parameters different modelling approaches are applied and compared – discriminant analysis, bivariate regression, neural and kriging models. For the parameter optimization aiming the improvement of the quality performance characteristics according to pre-set technological requirements multicriteria robust optimization approach is applied. The proposed approach is based on overall robust desirability function, consisting of two components: characteristic generalized desirability function and variance generalized desirability function. They can be weighted differently according to the importance of each of the components for the specific cases. Weights, which depend on the accuracy of the estimated models, are used for the individual desirability functions.

Keywords: Robust Models, Multicriteria Optimization, Bivariate Regression, Discriminant Analysis, Kriging Models, Neural Models, Electron Beam Welding.

Acknowledgements: The research was conducted within the framework of the project KP-06-N27/18, funded by the National Fund “Scientific Investigations”.

An Asymptotic Minimax Risk Description for the Exponential Two-Armed Bandit

Alexander Kolnogorov[1], Denis Grunev[2]
[1,2]Yaroslav-the-Wise Novgorod State University, Velikiy Novgorod, Russian Federation

We consider an asymptotic behavior of the minimax risk of the two-armed bandit as the control horizon N grows infinitely. Incomes of the two-armed bandit are exponentially distributed with parameters that are fixed, are a priori unknown to the person controlling the process, and depend only on currently chosen one of two possible actions. One has to determine the action corresponding to the largest expected one-step income and to ensure its predominant usage. According to the main theorem of the game theory, we search for the minimax risk as the Bayesian one computed with respect to the worst-case prior distribution. We show that maximal values of the regret are attained in the domain of “close” distributions where mathematical expectations of one-step incomes differ by the magnitude of the order of N-1/2. Then we consider batch processing when the same actions are applied to batches of incomes and then cumulative incomes in batches are used for the control. If the sizes of batches are large enough then corresponding minimax risk is equivalent to the minimax risk of the Gaussian two-armed bandit with properly assigned parameters of one-step incomes. Finally, we consider asymptotic description of the Bayesian risk of the exponential two-armed armed bandit which processes incomes one by one. In the domain of “close” distributions, this Bayesian risk is described by the same second order partial differential equation as the Bayesian risk of the Gaussian two-armed bandit. This means that optimal one-by-one processing of incomes of the exponential two-armed bandit is not more efficient that optimal batch processing as N grows infinitely.
The reported study was funded by RFBR, project number 20-01-00062.
Keywords: exponential two-armed bandit, minimax approach, Bayesian approach, main theorem of the game theory, batch processing.

Discrete time convolution and application to Renewal and Markov Renewal Theory

Leonidas Kordalis[1], Samis Trevezas[1]
[1]Department of Mathematics, University of Athens, Athens, Greece

In this study, we focus on some foundational aspects of discrete time convolution of real and matrix valued functions. Barbu and Limnios [Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their Use in Reliability and DNA Analysis; 2008] used this framework to solve Markov Renewal equations and study properties of characteristics related to semi-Markov chains. In this work, inspired by the aforementioned work, we revisit discrete time convolution, we develop an appropriate algebraic formalism and we highlight the use of convolution in some well-known results in combinatorics, the theory of discrete random variables, Renewal and Markov Renewal theory. Special mention is given to the convolutional inverse which plays a fundamental role to obtain unique solutions for Markov renewal equations.
Keywords: discrete time convolution, algebraic properties, renewal theory, Markov renewal theory.

Monte-Carlo Accuracy Evaluation of Laser Cutting Machines

Samuel Kosolapov
Department of Electronics, ORT Braude Academic College of Engineering, Karmiel, Israel

A big selection of laser cutting machines is available now – from 100 USD entry level devices to 50,000 USD industrial machines. Generally, accuracy of the specific model is characterized by a simple numerical parameter – from 0.3 mm for simple models to 0.01 mm for industrial models. However, using one parameter to characterize engraving accuracy in some situations may be misleading. This single parameter may be adequate to evaluate the accuracy of, say, horizontal cut, but more parameters may be required to adequately describe accuracy of the cut between two arbitrary points. In order to evaluate the practical accuracy of the different mechanical designs of the laser cutting machines, MAPLE

based software simulator was designed. By changing type of the mechanical design and values of parameters of geometrical sizes of the mechanical members used, mechanical slacks and mechanical rigidity, practical evaluation of the resulting cut accuracy for different parts of the cut can be calculated. MAPLE models for the following mechanical designs were tested: 2D and 3D models using stepper motors with timing rubber belts, 2D and 3D models using stepper motors with lead screws, 2D two-arm model using servo motors, 2D Pintograph-based design with high resolution servo motors. To make simulation practically useful, all parameters are characterized by their ranges, so that Monte-Carlo approach can be used to get a statistical evaluation of the resulting accuracy. Simulation' results are presented as a 2D maps, so that the accuracy of cut at different points can be easily seen. By using those 2D maps a customer may evaluate if the accuracy of the selected model is adequate for the specific task. Simulation results show that inexpensive Pintograph-based models driven by high definition PWM servo motors may be adequate for most non-industrial hobby customers.

Keywords: Monte-Carlo Simulation, Accuracy Evaluation, Laser Engraver, Laser Cutting Machine, MAPLE Model, Pintograph.

Using Parameters of Piecewise Approximation by Exponents for Epidemiological Time Series Data Analysis

Samuel Kosolapov

Department of Electronics, ORT Braude Academic College of Engineering, Karmiel, Israel

Our days a number of detailed epidemiological data are available in the form of time series data: N(Ti) – where N is documented number of events registered at an equidistant time moments Ti (for example: "Number of newly reported cases for COVID-19 in the last 24 hours" - published on the daily basis by WHO). Theoretically, those data can be adequately described by a different dynamic models containing exponential growth and exponential decay elements. Practically, parameters of those models are not constants – they can change in time because of many factors like changing hygiene policies, changing social behavior and vaccination.

Hence, it was decided to use piecewise approach: short sequential fragments of time series data are approximated by a function, containing some parameters. The above parameters are evaluated for the first time series data fragment. Then, next data fragments are processed. As a result, new time series data are created: evaluated sequences of parameters. Those new series can be considered and analyzed as a function of time. In the simplest example function to be used for every fragment is: A + B*exp(t / C). Resulted values of A, B, C in that case are time series data A(Ti), B(Ti) C(Ti) known at the equidistant time moments Ti. Their values can be as positive, as negative. By plotting those sequences, it can be seen if simple growth or decay model is adequate. Significant jumps of values may point to an interesting event – for example – to the start of massive vaccination or to the effect of a non-desirable social behavior at the specific date. In order to make calculations robust, some preliminary filtration and after-filtration can be used (for example, by using Gaussian filter, Bessel filter, etc.). For a big number of data, models with more parameters can be used. For example: A + B*exp(t/C) + D*exp(t/E). This model may be more adequate when decay process (described by negative constant E) has some time delay relative to the growth process (described by positive constant C). A number of practical examples were considered.

Keywords: Time Series Data, Approximation by Exponents, Epidemiological Data Analysis.

On Entropy-type measures and divergences with applications

Ch. Koukoumis, A. Karagrigoriou

Lab of Statistics and Actuarial-Financial Mathematics University of the Aegean, Greece

Information theory is a branch of pure and applied sciences that deals with the quantification of information. It started as being a key player in modern communication theory by formulating a communication system as a stochastic process. Tuller (1950) initially and Pierce (1956) later observed that the strong similarities between the underlying mechanisms of communication theory and information theory. The evolution of the field as well as the mathematical rigor that governs it are attributed to three

great researchers, namely Fisher (1956), Shannon (1956) and Wiener (1956). The most fundamental measure in information theory is entropy which was first recognized, formulated and defined in statistical mechanics (Fisher (1936); Shannon and Weaver (1949)) and consequently triggered the enormous development of the field, in the years that follow. In this work we review entropy-type measures and divergences, discuss their properties and unfold their diverse applicability. In should be noted that the concept of entropy was used firstly in Physics, and more specifically in the field of thermodynamics (Clausius, 1865) while its statistical definition was developed by Boltzmann (around 1870) but its applications go beyond Physics. In the present work we attempt to approach the entropy from a probabilistic or stochastic viewpoint and combine it with the concept of distance which can find numerous applications in Applied, Engineering or Management Sciences. The concept of divergence is fundamental in data analysis since it quantifies the distance between two populations, two models or two functions. By combining the two concepts and relying on entropic-type divergences or measures we could provide both researchers and practitioners with useful probabilistic tools for modelling purposes in various scientific areas including goodness of fit in reliability theory or survival analysis, portfolio selection in financial mathematics, decision making in management sciences etc.

Keywords: Conference, ASMDA, Entropy, Relative entropy, Divergence measures, Earthquakes.

Random Walk with Varying Transition Probabilities Applied on Tennis Modelling

Tomáš Kouřim[1], Petr Volf[2]

[1]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic, [2]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague

Possible application of the recently introduced [1] random walk with varying transition probabilities on professional tennis modelling is investigated in the present paper. The model considers a one-dimensional random walk with discrete time, steps +-1 and the position of the walker

controlled by history dependent transition probabilities. The model is applied on a large dataset consisting of professional tennis results from seasons 2010 through 2020. The dataset is divided into different subsets based on gender, surface, season, tournament type or even just a single player or group of players. Model's performance is measured on the entire dataset as well as on all available subsets. The prediction power is studied both from the statistical point of view as well as compared to real life odds provided by a bookmaker. The model's accuracy is further compared with that of a logistic regression model. Possible improvements of the model are briefly introduced and its usage in real life setting for betting against a bookmaker are discussed.

References:

[1] Tomáš Kouřim and Petr Volf. Discrete random processes with memory: Models and applications. Applications of Mathematics, volume 65, pages 271-286. Institute of Mathematics, Czech Academy of Sciences, 2020.

Keywords: Random walk, history dependent transition probabilities, tennis modelling, sports betting.

The correlation between oxygen consumption and excretion of carbon dioxide in the human respiratory cycle

Anatoly Kovalenko[1], Konstantin Lebedinski[2], Verangelina Moloshneva[3]

[1]Ioffe Institute of Russian Academy Science, St. Petersburg, Russia,
[2]Mechnikov North-western State Medical University, St. Petersburg, Russia,
[3]ITMO University, St. Petersburg, Russia

Treatment of lung damage in coronavirus infection and other pathologies with the use of artificial ventilation, surgery under general anesthesia require strictly monitoring of the consumption and transport of O2, as well as the excretion of CO2 in the patient's respiratory cycle. Oxygen transport can be easily monitored in real time by means of percutaneously pulse oximetery based on the level of arterial oxygen saturation. The method is based on the changing the wavelength of light absorbed by blood in the red and infrared parts of the spectrum, dependent on the number of oxygen molecules related by hemoglobin. The control of the O2 content in the inhaled and exhaled air, and the respiratory removal of CO2 can be carried out using inertial mechanical gas analyzers based on the

separation of a component from the gas mixture by special absorbers and measuring changes in the sample volume at constant pressure, or pressure at a constant volume of the measuring chamber. Electroc hemical, acoustic, thermal, magnetic, ionization and other types of gas analyzers based on the change in the relevant properties of the measured gases, depending on their concentration in the mixture, are also used. Operations research of synchronous data of current monitoring: indicators of air flow velocity in the patient's respiratory cycle and time capnogram for measuring the partial pressure of CO2 released by the body in the exhaled air mixture, are more adequate to the clinical conditions of intensive care units. The report gives an algorithm and the results of such an analysis, which also take into account air humidification in the respiratory tract.
Keywords: gas analysis, respiratory cycle, correlation O2 and CO2.

Application of an SEIRD mathematical model to the COVID-19 pandemic

Ashok Krishnamurthy[1], Loren Cobb[2], Bedrich Sousedik[3], Maya Mueller[3], Agatha Ojimelukwe[4]

[1]Department of Mathematics and Computing, Mount Royal University, Calgary, Alberta, Canada, [2]University of Colorado Denver, CO, USA, [3]University of Maryland Baltimore County, MD, USA, [4]University of Port Harcourt, Rivers, Nigeria

The global coronavirus pandemic (COVID-19) reached Lagos, Nigeria on February 27, 2020. Since then, COVID-19 infections have been reported in the majority of Nigerian states. We present a stochastic Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) compartmental model of epidemiology to capture the spatiotemporal transmission dynamics of the spread of COVID-19 and provide insight that would support public health officials towards informed, data-driven decision making. Data assimilation is a general category of statistical tracking techniques that incorporate and adapt to real-time data as they arrive by sequential statistical estimation. Data assimilation applied to the SEIRD model receives aggregated epidemiological data from the Nigeria Centre for Disease Control (NCDC) and uses this data to perform corrections to the current state vector of the

epidemic. In other words, it enhances the operation of the SEIRD model by periodically executing a Bayesian correction to the state vector, in a way that is, at least arguably, statistically optimal. We observe that the prediction improves as data is assimilated over time. It is essential to understand what future epidemic trends will be, as well as the effectiveness and potential impact of government disease intervention measures. Predictions for disease prevalence with and without mitigation efforts are presented via spatiotemporal disease maps.

Keywords: Stochastic Models, Epidemiology, Data Assimilation, Bayesian Statistics, COVID-19.

Nonlinear connectivity as a driver of time-horizon heterogeneity

Catherine Kyrtsou[1,2]*, Christina Mikropoulou[3], Angeliki Papana[4]

[1]University of Macedonia, Department of Economics, [2]CAC, IXXI Lyon and EconomiX University of Paris 10, [3]University of Macedonia, Department of Economics, [4]University of Macedonia, Department of Economics

The effect of sample size on the performance of causality measures constitutes an appealing issue when researchers look for causal inference in simulated and real time series. The intensity and the nature of the relationship among variables affect significantly the informational content in small and large data samples. Real high-dimensional systems hide strong interdependences that can arise not only because of direct causal effects but also because of autocorrelation effects within each time series, indirect effects, or common drivers. In a recent joint paper, we highlight the tight linkage between the degree of nonlinearity and the detection of false positive causal couplings in various multivariate systems. We further elucidate this behavior by using financial data in which time horizon seems to crucially determine causal directionality due to the activation of heterogenous trading activity. In the present study, we provide additional evidence that complex connectivity in longer samples helps direct causality measures to identify true positive couplings but at the same time it contributes to the appearance of many false positive cases as the coupling strength intensifies. In the simulation experiment we use both linear and nonlinear multivariate systems, as well as fully and partially conditioned causality measures.

On the Bivariate Non-central Polya-Aeppli distribution and application

Meglena D. Lazarova[1], Leda D. Minkova[2]
[1]Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Sofia, Bulgaria, [2]Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria

In this paper, by the trivariate reduction method we introduce a bivariate Non-central Polya-Aeppli distribution (BNPAD). We discuss a number of properties of this distribution including the probability generating function, correlation structure, probability mass function and recursion formulas. As application we consider the bivariate risk model with BNPA distributed claim number. The ruin probability is also discussed.
Keywords: Non-central Polya-Aeppli distribution, trivariate reduction method, Bivariate Non-central Polya-Aeppli distribution, bivariate risk model
Acknowledgements: The authors are totally supported by project DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

Stability of Unreliable Jackson Networks and Approaches to Bottlenecks Recognition

Elena Lenena
Department of Mathematics and Mechanics, Moscow State University, Moscow, Russia

We consider a generalized Jackson (open queueing) network with regenerative input flows. The system is unreliable, so that every server is a subject to a random environment influence generating breakdowns and repair. Stability conditions of such networks are investigated based on the results for reliable networks and additional considerations. Special attention is paid to the behavior of each server specifically, not just the whole system.
Applying the theorem on the strong approximation of the vector of queue lengths and the local averaging techniques, we construct consistent

estimates of limit process parameters. Then, an approach to bottlenecks recognition is presented. The aim of this statistical inference problem is to detect the buffers in which the queue length will systematically approach infinity, as well as those for which it is stochastically bounded. We assume that for each station, the general input flow (cumulated input streams from outside of the system and streams being transferred from other stations) and queue length vector are known. The unreliable server concept implies that the breakout and repair times also being observable. As for the routing matrix, we assume that it is an attribute of the system and either is known or estimated based on other data. Using the estimates, we provide an algorithm for the identification of strict bottlenecks and non-bottlenecks.
Keywords: Jackson networks, Stability conditions, Unreliable systems, Bottlenecks recognition.

Investigating the impact of causes of death on longevity improvements in Italy

Susanna Levantesi[1], Andrea Nigri[2], Gabriella Piscopo[3]
[1]Department of Statistics, Sapienza University of Rome, Italy, [2]Department of Agricultural Sciences, Food, Natural Resources and Engineering, University of Foggia, Italy, [3]Department of Economics and Statistical Science, University of Naples Federico II, Italy

Since the last two centuries, longevity evolution has significantly influenced the population dynamics, showing a rapid decline in mortality rates, driven by medical progress and better living conditions. As a consequence, life expectancy at birth has notably increased over time, without any sign of an impending limit. Overall, life expectancy at birth has exhibit a constant pace, following an almost homogeneous shape with different country-specific transition phases. Life expectancy is indeed an index that incorporates the variation of length of life (or lifespan inequality) and conceals country-specific phases of evolution led by heterogeneity behaviour. After the IIWW, mortality improvements have been driven by a shift of the modal age, accompanied by compression of the age at death distribution around the mode. In light of that, countries that exhibit a remarkable transition phase in life expectancy, show a negative correlated behaviour in lifespan inequality. Our paper contributes to the literature by evaluating the contribution of age-specific cause of death mortality to the

changes in life expectancy and lifespan inequality. We use decomposition techniques, which enable to impute the age and causes of death accountable for changes in life expectancy or lifespan inequality between two specified periods.
In this way, we can explain the changes of the transition phases that Italy has shown after the IIWW, providing the causes of death contribution by age in reducing lifespan inequality, translating them into gains in life expectancy.
Keywords: life expectancy, causes of death, heterogeneity, longevity.

Predicting the second wave of COVID-19 pandemic through the Dynamic Evolving Neuro Fuzzy Inference System

Susanna Levantesi[1], Andrea Nigri[2], Gabriella Piscopo[3]
[1]Department of Statistics, Sapienza University of Rome, Rome, Italy
[2]Department of Agricultural sciences, Food, Natural resources and Engineering, University of Foggia, Foggia, Italy, [3]Department of Economic and Statistical Science, University of Naples Federico II, Naples, Italy

In this paper we propose to implement the methodological framework of the Dynamic Evolutionary Neuro Fuzzy Inference System to offer an analysis of the possible evolution of the second wave of COVID-19. The choice of the model is motivated by the fact that the spread of the pandemic must be read in its dynamism and every prediction cannot ignore the daily updating of available data and new information. We provide results of the prediction of the second wave of COVID-19 across Europe and USA, soliciting to update the model day by day as new information occurs. We show that the Dynamic and Evolutionary Neuro Fuzzy Inference System produces better results than previously published fuzzy models applied to the predictivity of the Covid spread. The study offers a tool to analyze the effectiveness of the virus containment measures in the short run for public health providers and the government in the battle for the control the COVID-19 spread.
Keywords: Adaptive Algorithm, Covid-19, Fuzzy Systems, Predictive Methods.

Life Expectancy and Different Parameter Identification in Chinese Retirement Plan

Yichao Li[1], Nan Li[2], Hong Mi[3*]

*[1]School of Public Affairs, Zijin'gang Campus, Zhejiang University, HangZhou, China, [2]Population Division, Department for Economic and Social Affairs, United Nations, 10017 New York, USA, 3*Corresponding Author, School of Public Affairs, Zijin'gang Campus, Zhejiang University, HangZhou, China*

Under the current statutory retirement age of 60 for male employees and 55 for female employees, pension annuity divisors in China is 139 months. However, this number was set in 2005, which cannot meet the background of life expectancy improvement in China, which will bring natural pressure to the operation of basic pension insurance for employees. In this paper, by constructing the life table of the elderly aged 65+ and Age-Augment-Log-Quad model, the elderly life expectancy and mortality rate aged 65-80(15q65) were calculated in China and its provinces in 2015, and the provincial spatial distribution characteristics and influencing factors of the elderly life expectancy were analyzed. The correlation between the elderly life expectancy and the pension gap were calculated. The results show that: (1) There is a large spatial imbalance in the elderly life expectancy aged 65+ among provinces in China; (2) Elderly life expectancy (65+) was over 17 years for men and over 19 years for women in China in 2015, which means that the pension annuity divisors in China should be promoted to over 200 months under the background of China's Gradual Delayed Retirement Plan in the future. The calculation results provide a theoretical basis from the perspective of mortality change for optimizing the setting of the annuity divisors for the statutory individual account of basic pension insurance, and provide a numerical reference for the formulation of China's Gradual Delayed Retirement Plan and the promotion of the National Pooling of pension insurance accounts.

Keywords: Chinese Retirement Plan; Elderly Life Expectancy (65+), Mortality Rate (15q65), Spatial Distribution, Annuity Divisors.

Interpreting a Topological Measure of Complexity for Decision Boundaries

Ian Lim[1], Michael Moy[2], Alan Hylton[3,] Robert Short[4]
[1]The University of Texas at Arlington, Arlington, USA, [2]Colorado State University, Colorado USA, [3,4]NASA Glenn Research Center, Cleveland, USA

We propose a method to examine the decision boundary of a classification algorithm to yield insight into the nature of overfitting. In machine learning, model evaluation can be performed via two common techniques: train-test split or cross validation. In this paper, we grow this toolkit to include tools from the field of topological data analysis. In particular, we use persistent homology, which roughly characterizes the shape of a data set. Our method focuses on binary classification, using training data to sample points on the decision boundary of the feature space. We then calculate the persistent homology of this sample and compute metrics to quantify the complexity of the decision boundary. Our experiments with data sets in various dimensions suggest that in certain cases, our measures of complexity are correlated with a model's ability to generalize to unseen data. We hope that refining this method will lead to a better understanding of overfitting and a means to compare models.
Keywords: Topological Data Analysis, Machine Learning, Data Science.

Bayesian Estimation to explain sugar beet growth

Dimitris Logothetis[1], Sotiria Malefaki[2], Samis Trevezas[1], Paul-Henry Cournède[3]
[1]Department of Mathematics, University of Athens, Athens, Greece, [2]Department of Mechanical Engineering & Aeronautics, University of Patras, Patras, Greece, [3]MICS laboratory, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

Plant growth modeling and estimation is a topic of increasing interest due to its potential applications. The GreenLab functional-structural plant growth model mimics the dynamics of plant growth particularly well. The GreenLab-1 model focuses on plants with deterministic organogenesis and it is particularly adapted for the growth of the sugar-beet plant. In this

study, we explore the benefits of Bayesian Estimation of the parameters of the model, including both functional and noise parameters. The case study concerns a real dataset from the sugar-beet plant. The vector of observations consists of organ masses measured only once at a given observed time, including blades, petioles, and the root of the sugar-beet plant. For the Bayesian Estimation, an appropriate Markov Chain Monte Carlo algorithm is developed, providing a useful tool for analyzing this type of models. The Bayesian approach has much more flexibility in handling complex structures, thus providing a useful tool for analyzing this type of models. Several implementation issues are also discussed.
Keywords: Bayesian estimation, MCMC, GreenLab model, sugar-beet plant.

A non-parametric Hawkes process model for road accidents

Pierfrancesco Alaimo Di Loro[1], Kieran Kalair[2], Colm Connaughton[2,3]

[1]Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy, [2]Centre for Complexity Science, University of Warwick, Coventry, UK, [3]Centre for Complexity Science, University of Warwick, Coventry, UK

We propose a self-exciting spatiotemporal point process as a model for incident data to verify the cascading effect of road accidents and model the rates of primary and secondary accidents. This process uses a background component to represent primary accidents, and a self-exciting component to represent secondary accidents. The background consists of periodic daily and weekly components, a spatial component and a long-term trend. The self-exciting components are decaying functions of space and time. These components are determined via kernel smoothing and likelihood estimation. The model is applied to data from the Rome urban road network and the UK National Traffic Information Service (specifically on the M25 motorway). The temporal window spans 5 months in 2017 for the first application and a 12-months period during 2017–2018 for the second. Results show that the self-excitation accounts for 5–7% of the events in both cases, with associated time and length scales of around 100-500 min and 500-1000 m, respectively. Temporally, the background is stable across seasons with a daily double peak structure reflecting commuting patterns. Spatially, the model was able to detect areas

infamously known as road accident hazards. In-sample and out-of-sample validation are performed to assess the model fit.
Keywords: Hawes process, Non-parametric statistics, Kernel density estimation, Space-time.

Bayesian hierarchical modeling for physical activity trajectories using Actigraph data

Pierfrancesco Alaimo Di Loro[1], Marco Mingione[1,2], Jonah Lipsitt[3], Christina M. Batteate[4], Michael Jerrett[3], Sudipto Banerjee[5]

[1]Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy, [2]Institute of Applied Computing "M. Picone" (IAC - CNR), Rome, Italy [3]Environmental Health Sciences Department, University of California, Los Angeles (UCLA) - Fielding School of Public Health, USA, [4]Center for Occupational & Environmental Health, University of California, Los Angeles (UCLA) - Fielding School of Public Health, USA, [5]Department of Biostatistics, University of California, Los Angeles (UCLA) - Fielding School of Public Health, USA

Rapid technological developments in accelerometers have generated substantial interest in monitoring human activity. Wearable devices, such as wrist-worn sensors that monitor gross motor activity (actigraph units) continuously record the activity levels of a subject, producing massive amounts of high-resolution measurements. Analyzing actigraph data needs to account for spatial and temporal information on trajectories or paths traversed by subjects wearing such devices. Inferential objectives include estimating a subject's physical activity levels along a given trajectory; identifying trajectories that are more likely to produce higher levels of physical activity for a given subject; and predicting expected levels of physical activity in any proposed new trajectory for a given set of health attributes. We devise a Bayesian hierarchical modeling framework for spatial-temporal actigraphy data to deliver fully model-based inference on trajectories while accounting for subject-level health attributes and spatial-temporal dependencies. We undertake a comprehensive analysis of an original dataset from the Physical Activity through Sustainable Transport Approaches in Los Angeles (PASTA-LA) study to formally

ascertain spatial zones and trajectories exhibiting significantly higher levels of physical activity.
Keywords: Bayesian Hierarchical models, Physical activity, Spatio-temporal statistics, Gaussian processes.

Asymptotic results for certain first-passage times and areas of renewal processes

Claudio Macci[1], Barbara Pacchiarotti[2]
[1,2]Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, Roma, Italy

We consider the process {x-N(t)}, where x > 0 and {N(t)} is a renewal process with light-tailed distributed holding times. We are interested in the joint distribution of (T(x),A(x)) where T(x) is the first passage time of {x-N(t)} to reach zero or a negative value, and A(x) is the corresponding first passage area. Our aim is to prove asymptotic results
as x tends to infinity in the fashion of large (and moderate) deviations.
Keywords: large deviations, moderate deviations, joint distribution, renewal theory.

Analysis techniques of coupling maximum and minimum values between data sets based on mathematics measures and indicators

Konstantinos N. Makris, Ilia Vonta
National Technical University of Athens, School of Applied Mathematical and Physical Sciences, Zografou Campus, Zografou, Athens, Greece

This paper studies alternative coupling techniques for maximum and minimum values between data sets. More specifically, in this work it is checked whether the values of two or more sets of data have their maximum or minimum values at the same time points for the case that data are depended of time, or at the same frequencies for the case that

data are random variables. For this purpose it is used direct and indirect methods based on mathematics measures and indicators.

Keywords: Index MKN, Index μ[MKN], Time Series.

Bipartite Fuzzy Stochastic Differential Equations: Generalization of the Lipschitz Condition

Marek T. Malinowski

Department of Applied Mathematics, Cracow University of Technology, Kraków, Poland

In the communication we will consider stochastic differential equations with values in the space of fuzzy sets. This kind of stochastic equation may be suitable for modelling phenomena subjected to random factors and whose values are not precisely described numerically. The properties of fuzzy sets mean that, unlike the classic stochastic form of differential equations, it now makes sense to consider such equations in a certain bipartite form. The talk will present the results concerning a theorem on existence and uniqueness of solution, assuming that the coefficients of the equation satisfy a different condition than the classic Lipschitz condition.

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

Special session on Topological Data Analysis

My Ismail Mamouni

Department of Mathematics, CRMEF Rabat, Morocco

As emerging area in computational topology where research is centered on the development of methodologies to study the “shape” of point cloud data. Specifically, Topological Data Analysis (TDA) offers a principled approach for the analysis of high-dimensional large volume data sets through the methods of topological compression and persistence

homology. As of late, TDA has found applications in bioinformatics, remote sensing, data mining, and computer vision, to name a few. This special session will serve as a venue for researchers to publish innovative state-of-the-art hybrid TDA/ML based learning systems for unsupervised and semi-supervised learning. The goal is to develop novel approaches for the extraction of topological features/summaries of data that can be exploited in learning systems to enhance performance and provide mechanisms to model system interpretability. TDA is a topic which has recently seen many applications. The goal of this special session is to highlight the bridge between applied statistic models and topological data analysis. In this way, we hope to encourage more engineers to start exploring TDA and its applications. This papers should briefly introduces the standard techniques used in this area, delineates the common theme connecting the works presented in this session, and concludes with a brief summary of each of the papers presented. Submitted papers are expected to present novel research with a practical study or novel application based on topological data analysis or a survey of the literature. This special session invites submissions on recent advances, approaches, theories and applications related to topological data analysis.

Weighted Hill Estimators for the extreme value index

Ayana Mateus[1], Frederico Caeiro[2]

[1,2]Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia and Centro de Matemática e Aplicações (CMA), Caparica, Portugal

In this work, we deal with statistics of univariate extremes and propose a new class of weighted Hill estimators for the extreme value index of Pareto-type models. The extreme value index dominates the tail behaviour and needs to be estimated in a precise way because other tail parameters such as an extreme quantile or a tail probability depends on the value of this parameter. We also derive their non-degenerate asymptotic distribution. Monte Carlo simulations were performed to analyse the finite sample behaviour of the proposed new estimators. A comparison with other important extreme value index estimators from the literature is also provided.

Keywords: Asymptotic distribution, Extreme Value Index, Semiparametric estimation.
Acknowledgment: "This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications)".

On the asymptotic power of the new test for equality of two distributions

Viatcheslav Melas[1], Dmitrii Salnikov[2]
[1]Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia, [2]Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia

A new test for equality of two distributions in a class of models was introduced in the paper [1]. It included normed sums of terms of the form: g ($X_i - {}^{Y}{}_j$), g ($X_i - X_l$), g ($Y_k - Yj$), i,l=1,…, n, j,k=1,…,m, where (X_1,…, X_n), (Y_1,…, Y_m) are two samples under consideration and g(x-y)= ln(1+(x-y)2). Earlier a similar test based on g(x-y) = ln |x–y| was introduced in [2] but no analytical results on asymptotic power were obtained there. No such results for that test appeared in other works to the best knowledge of the authors. Here the distributions are assumed to be such that the expectation of ln (1+u2), where u is the corresponding random variable, is finite. In [1] it was proved that the distribution of the test statistics tends under natural conditions to that of (aZ+b)2, where Z is the random variable with the standard normal distribution, and a and b are some parameters to be estimated numerically. Here we discuss an appropriate choice of estimates for a, and b, and the corresponding estimates for asymptotic power. Simulation results confirm that the approach allows to evaluate the power of the test with a high accuracy under moderate numbers n and m for the case where the compared distributions are normal or Cauchy. Also, we discuss some conjectures about optimality properties of the test considered. The work was supported by RFBR (grant N 20-01-00096).

References:

[1] Melas, V., Salnikov, D. (2020). New test for equality of two disributions. In: Kozyrev D. (Ed.), THE 5th INTERNATIONAL CONFERENCE ON STOCHASTIC METHODS (ICSM-5): Proceedings of the international scientific conference, pp. 125-129.
[2] Zech, G., Aslan, B. (2005). New test for the multivariate two-sample problem based on the concept of minimum energy. Journal of Statistical Computation and Simulation 75(2), pp. 109-119.
Keywords: Test for equality of two distributions, Asymptotic power.

Weather index-based insurance in advanced agricultural risk management

Massimiliano Menzietti[1], Marco Pirra[2]
[1,2]Università della Calabria

The increase in the frequency and severity of extreme weather events (droughts, storms and others) associated with climate changes has relevant impacts for European agriculture and reduces livelihood options for millions of small-scale farmers in low-income countries. Weather index-based insurance schemes are a possible option in managing weather and climate risks because they refer to weather station data or grid data (rainfall, temperature) in order to assess the payouts and these are therefore immediate and with fewer disputes than conventional crop insurance coverages. These insurance schemes are at a developmental stage but can be considered an attractive opportunity, due to their advantages: low costs, no information asymmetry, abundant data, wide spectrum of activities covered, flexibility. Though, one of the critical aspects is the susceptibility to basis risk, mainly the spatial one: the reduction of this risk is crucial to exploit the opportunities offered by these schemes. The goal of the study is to investigate the potential benefits machine learning algorithms and predictive analytics techniques could offer in this area, reducing the underwriting inefficiencies, developing better insurance products and improving sustainable strategies to make agricultural sector more resilient.
Keywords: Insurance, machine learning, weather events, risk management.

Multiway Contingency Table Analysis based on a New General Class of Test Statistics

Christos Meselidis, Alex Karagrigoriou
Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, Samos, Greece

Measures of divergence are used extensively in statistics in various fields. Many of the currently used tests, such as the likelihood ratio, the chi-squared, the score and Wald tests, can in fact, be shown to be defined in terms of appropriate divergence measures. Tests like the above could be used for the analysis of multiway contingency tables for statistical inference. In this work we consider statistical inference for threedimension contingency tables based on a general class of measures, namely the (Φ, α)-power divergence family for multinomial populations. In particular, for the purpose of estimating the unknown parameters which satisfy some functional relationships we exploit the restricted minimum (Φ, α)-power divergence estimator and then use it in the new double index (Φ, α)-test statistic for the problem of goodness-of-fit. Through an extensive simulation study, we compare in terms of size and power the behavior of the proposed tests along with the classical ones.
Keywords: Divergence Measures, Multivariate Data, Goodness-of-fit Tests, Multiway Contingency Tables.

A Spatial functional spectral approach to COVID-19 incidence in Spanish Communities

Doris Miranda[1], M. D. Ruiz[1]

[1]Department of Statistics and Operational Research. University of Granada, Spain

In this paper, we adopt a pure point and continuous spectral approaches, for predicting COVID-19 incidence from a Bayesian and a nonparametric frameworks, respectively. Specifically, in the first case, we consider a particular example of the dynamical multiple linear regression model in

function spaces, introduced in Ruiz--Medina, Miranda and Espejo (2019), under diagonal regressors. The functional regression parameter vector is estimated in terms of the Bayesian approximation of the entries of the sequence of inverse matrixes, involved in the definition of the computed Hilbert -- valued generalized least--squares estimator (see equation (24) in Ruiz--Medina, Miranda and Espejo (2019)). Under this functional linear modeling, spatial correlations are reflected in the matrix covariance operator of the functional error term. On the other hand, in the second case, we adopt a continuous spectral approach, assuming spatial stationarity in the functional correlation model, representing possible interactions between the COVID-19 incidence curves at the Spanish Communities analyzed. We reformulate, for spatially distributed correlated curves, the nonparametric estimator of the spectral density operator, based on the functional periodogram, proposed in Panaretos and Tavakoli (2013) in the functional time series context. Finally, a comparative study is carried out to assess the performance of both approaches in the prediction of COVID-19 incidence.

Keywords: Spatial functional data analysis, spectral analysis, Covid-19.

Network Failure Problem of Random K-out-of-N Systems

Hiroaki Mohri[1], Jun-ichi Taeshita[2]

[1]Faculty of Commerce, Waseda University, Shinjuku-ku, Tokyo, Japan,
[2]Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

The idea of “Random K-out-of-N System” was proposed by Ito and Nakagawa [1] in the reliability theory literature. The system does work when at least K units out of N units do not break down. Here, K is a random variable. We applied this idea to Network Failure Problem. Suppose undirected simple graphs without parallel edges, G(V,E). Units in Reliability Theory correspond to edges of the graphs. We suppose only edges may fail on the graphs. For nodes, they never fail. If a connected graph become an unconnected graph by failures of edges, we call it “Network failure”. To get probability distribution of $P(K < k)$ for the graphs, we use recurrent relations. We show our algorithm by 2 types of graph generation procedures.

References:
[1] K. Ito and T. Nakagawa, “Reliability properties of K-out-of-N: G systems” in Systems Engineering: Reliability Analysis Using k-out-of-n Structures, M. Ram and T. Dohi, Eds, Boca Raton, FL: CRC Press, 2019, pp.25-40, ISBN: 978-1-138-48292-0.
Keywords: Reliability, k-out-of-n system, Graph, Network, Applied Probability.

Decomposing Differences in Life Expectancy with and without Disability: The Case of Czechia

David Morávek[1], Jitka Langhamrová[2]
[1,2]Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Demography, Prague, Czech Republic

The improvement in mortality is reflected in the fact that people have a longer life expectancy. At the same time, healthy life expectancy has been increasing. Nevertheless, the growth in life expectancy and healthy life expectancy is not exactly the same. What is the part of life that people live on average in a good health without a disability? The aim of this paper is to evaluate life expectancy at birth and disability-free life expectancy in Czechia for males and females over the time. The next part of the evaluation of the mortality and the health level of the Czech population deals with decomposing differences in healthy life expectancy into the additive contribution of the mortality vs. the effect of a disability, of different age groups and of different causes of death. This decomposition method was originally developed by Nusselder & Looman (2004) with regard to the Sullivan method that calculates healthy life expectancy and is an extension of the Arriaga method (1984) that decomposes differences in the overall life expectancy. The decomposition tools are available for performing calculations as part of the EHEMU project (2007–2010).
Keywords: Life expectancy, disability-free life expectancy, Sullivan method, decomposition method.

Mother and child health: Mal-nutrition, Anemia, Food habits along with MMR, MCH making India the worst position in Hunger index

Barun Kumar Mukhopadhyay
Population Studies Unit, Indian Statistical Institute, India (Retired scientist)

Very recently world news circulated that India ranked 102 among 117 countries in the GHI (Global Hunger Index) 2020. The report has placed India's 94th position among 107 countries, much behind Bangladesh, Pakistan and Nepal. Further India has the highest prevalence of wasted children under five years in the world, which reflects acute undernutrition, according to the above index in 2020. The situation has worsened in the 2015-19 period, when the prevalence of child wasting was 17.3 per cent in comparison to 2010-14, when it was 15.1 per cent. Historically there were similar types of comments and criticisms about India. In 2012 the International Food Policy Research Institute (IFPRI), USA while prescribing their 1,000 day window mentioning the MCH condition of India was very precarious. The scientists and experts of the IFPRI were of the opinion and agreed that improving nutrition during the critical 1,000 day window is one of the best investments to achieve lasting progress in health and development of child and mother of India. The Global Nutrition Report 2017 (the first report published in the month of November, 2017) presented at Milan, Italy clearly gives a grim nutritional status of Indian mothers with 51 per cent suffering from anemia and 22 percent overweight. About MMR, India was to achieve a goal of 109 maternal deaths due to child birth and puerperum per 100,000 live births (MMR) by 2015 as par the guidelines of the Millennium Development Goal (MDG) of the United Nations (UN) and the other international bodies in the year 2000. Next time SDG (Sustainable Development Goals) was adopted at a mega conference in Rio de Janeiro in 2012 as the MDG in 2000 was not at all achieved. At present there are global worst situation as far as present subject of MCH due to the Covid-19, where Indian reports are pouring in. Present study tries to cover up most of the subject matter in order to understand the overall situation of India. The data are mostly government reports and different websites. The final results are yet to come.
Keywords: GHI, Covid-19, MCH, IFPRI.

FANOVA applied to mortality from Covid-19 in Colombia

Diana Paola Ovalle Muñoz
Department of Statistics and Operations Research of the Universidad de Granada, Granada, Spain

Functional data methodology is a branch of statistics that has recently been applied to different fields of knowledge. Its main utility is that it allows modeling the behavior of a continuous phenomenon through a function, which is treated as an observation, thus, a set of functions can be analyzed using functional statistics. On the other hand, in the last year, the Covid-19 phenomenon has aroused analytical interest, especially the mortality rate, which is an indicator of the severity of the disease and its effect on humanity, therefore, in this work it is proposed to model the behavior of the monthly mortality rate due to covid 19 in Colombia. The main objective is to compare the monthly mean curve of the death rate by departments using the Functional Analysis of Variance (FANOVA) proposed by Ramsay & Silverman (2005).
Keywords: FDA, FANOVA, Mortality from Covid 19.

Stochastic Optimal Control with a Multiscale Volatility Model

Jean-Paul Murara
School of Education, Culture and Communication, Malardalen University, Västerås, Sweden

In this paper, using generalised reference probability space we start by giving an introduction of a strong and also a week formulation of an optimal control problem. We extend the general problem to the particular stochastic optimal contol problem. After, we study the application of the above in the case of a portfolio optimization with a utility function that follows a multiscale stochastic volatilitity model.
Keywords: PDEs, SDEs, HJB equation, SV model, Porfolio optimization, Dynamic programming principle.

Dynamic Bayesian Network for modeling the Covid-19 epidemic

Federica Nicolussi[1], Ahmed Alsayed[1] Luisa Ferrari[2], Giuseppe Gerardi[1], Alessandra Micheletti[3], Giancarlo Manzi[1], Silvia Salini[1], Vincente Rios[1]

[1]Department of Economics, Management and QUantitative Methods, University of Milan, Milano, Italy, [2]University College London, UK, [3]Department of Environmental Science and Policy, University of Milan, Milano, Italy

In this work, the epidemic of Covid-19 in Italy is modeled with a dynamic model. We focused on the provincial (i.e. EU NUTS-3) level. Thus, official data from the Italian Ministry of Health are integrated with data extracted from daily official press conferences of regional authorities and from local newspaper websites. The different states of the disease are modeled with the dynamic Bayesian network for compartmental models in order to estimate the duration of time an individual is infectious, the heterogeneity in the risk of acquiring infection, and the patterns of seasonality (in the observed data). In particular, we incorporate in the latent infection state prevalence over time. Two periods, in correspondence with the two waves, are analyzed separately due to the different policies in the swab administration. Keywords: Covid-19, provincies, chain graphs, dynamic models.

References:

1. Anacleto, O., and Catriona Q. Dynamic chain graph models for time series network data. Bayesian Analysis 12.2 (2017): 491-509.

2. Ferrari, L., Gerardi, G., Manzi, G., Micheletti, A., Nicolussi, F., Salini, S. (2020). COVID-19 in Italy: An app for a province-based analysis, ArXiv 2020; Preprint: 1?13.

3. Enright, C. G., Madden, M. G., & Madden, N. Bayesian networks for mathematical models: techniques for automatic construction and efficient inference. International journal of approximate reasoning, 54(2), 323-342 (2013).

4. Ferrari, L., Gerardi, G., Manzi, G., Micheletti, A., Nicolussi, F., Salini, S. (2020). Modelling provincial Covid-19 epidemic data in Italy using an adjusted timedependent SIRD model, arXiv: 2005. 12170v2 .

5. Lauritzen, S. L. (1996). Graphical Models. Oxford University Press. MR1419991. 491, 493

6. Ozanne, M. V., Brown, G. D., Oleson, J. J., Lima, I. D., Queiroz, J. W., Jeronimo,S. M.,& Wilson, M. E. (2019). Bayesian compartmental model for an infectious disease with dynamic states of infection. Journal of applied statistics, 46(6), 1043-1065.

Multiple Trajectory Analysis in Finite Mixture Modeling

Cédric Noel[1], Jang Schiltz[1]
[1]University of Luxembourg Department of Finance, Luxembourg

Finite Mixture Models, as defined by D. Nagin (2005) and extended by J. Schiltz (2015) are used with normal, ZIP and logistic underlying distributions. D. Nagin (2005) extends his basic model to the dual trajectory analysis. But his method has the drawback that the many parameters in his model do not have a practical interpretation. We present another methodology which allows for multiple trajectory analysis with parameters that all have an easy interpretation. The model is illustrated by numerical examples and can be fitted with the R package trajeR developed by C. Noel (2021).
Keywords: Identifiability, Finite Mixture Models, ZIP distribution, logistic distribution.

Pricing Financial Derivatives in the Hull–White Model Using Cubature Methods on Wiener Space

Hossein Nohrouzian[1], Anatoliy Malyarenko[1], Ying Ni[1]
[1]Division of Mathematics and Physics, Mälardalen University, Västerås, Sweden

In our previous studies, we developed novel cubature methods on Wiener space of orders 5 and 7 in the sense that the cubature formula is exact for all multiple Stratonovich integrals up to a dimension equal to the order. In this paper, we apply the above methods to the modelling of fixed–income markets via affine models. Then, we apply the obtained results to pricing interest rate derivatives in the Hull–White model.
Keywords: Hull-White model, cubature method, Stratonovich integral, Wiener space.

Approximate Bayesian inference using mean-field distribution simulations

Antonin Della Noce[1], Paul-Henry Cournède[1]
[1]MICS laboratory, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

Dynamical systems representing populations of interacting heterogeneous individuals are rarely studied and validated within a Bayesian framework, with the notable exception of Schneider et al. (2006), dealing with a model of plants in competition for light resource. The reasons for this lack of coverage of a subject with such significant stakes (agriculture, crowd dynamics) are to be found in the computational difficulties posed by the problem of inference when the size of the population is large. In this paper, we focus on dynamical systems admitting a mean-field limit distribution when the population's size tends to infinity, such as the flocking models presented in Carrillo et al. (2010). We introduce a numerical scheme to simulate the mean-field distribution, which is a partial differential transport equation solution, and we use these simulations to simplify the likelihood distributions associated with Bayesian inference problems arising when the population is very partially observed.
Keywords: Approximate Bayesian inference, Mean-field limit, Numerical methods for partial differential equations.

A Comparison of Hidden Markov and Regression Models on Crop Stage Percentages Prediction

Ioannis Oikonomidis[1], Samis Trevezas[1]
[1]Department of Mathematics, University of Athens, Athens, Greece

Precision Agriculture (PA) has been receiving increasing interest in the last few years due to the rapid advancement of information technology. Among the many issues addressed by PA is the matter of crop stage percentages prediction with remote sensing techniques. In this study, a Hidden Markov and a Regression model are developed and compared with each other as well as the historic mean percentage predictions. The models mobilize two environmental characteristics commonly used in agricultural studies: The Accumulated Growing Degree Days (AGDD) and

the Normalized Difference Vegetation Index (NDVI). The performance of the models is illustrated with a case study on the Crop Progress Reports (CPR) of the US Department of Agriculture (USDA); particularly, the case study concerns the corn fields of Nebraska over a 10-year interval, from 2002 to 2011. The AGDD is calculated using data from meteorological stations across Nebraska and the NDVI is calculated using the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor. The initial results indicate that the regression model performs better on this large-scale study.

Keywords: Crop Stage Percentages Prediction, Hidden Markov Model, Regression Model, AGDD, NDVI.

Predicting Child Health Outcome in Ghana using Partial Least Squares (PLS) Structural Equation Modeling (SEM)

Frank Okwan

Department of Statistic and Demography, University of Szeged, Hungary

PLS-SEM is non- parametric statistical estimation method that has been proven both theoretically and empirical to produce good estimates even with small size. It is an estimation technique deemed appropriate for examining complex causal relationship between variables. The global under five mortality rate has decline from 93 per 1,000 live birth in 1990 to 41 in 2016 and subsequently to 38per 1000 live birth in 2019. The under-five mortality rate for Ghana, a country in sub-Saharan Africa is still a major public health challenge despite numerous initiatives implemented by government and other international stakeholders. The main objective of this paper is to proposed a causal model for predicting child health outcome measured by under-five mortality in Ghana. The results from PLS-SEM estimation showed both direct and effect on child health outcome measured by under-five mortality rate (U5MR). The results from the PLS-SEM estimation further showed that gross domestic product(GDP) has both direct and indirect effect on under five mortality rate(U5MR) a proxy for child health outcome, maternal mortality and out-of-pocket expenditure, immunization also have direct effect on child health outcome in Ghana. The effect of out-of- pocket expenditure is the greatest

and it is through immunization. The results indicate that immunization measured by Hepatitis B and Measles, out-of-pocket, GDP and maternal mortality are indicators that best predict child health outcome in Ghana. The proposed model showed important causal relationships needed for the implementation of interventions that will best address the poor child health outcome in Ghana.

Keywords: Causal model, Partial Least Squares, Under five mortality rate, Ghana.

Differences in the structure of infectious morbidity of the population during the first and second half of 2020

Vasilii Orel[1], Olga Nosyreva[1], Tatiana Buldakova1, Natalya Gureva[1], Viktoria Smirnova[1]

[1]Saint-Petersburg State Pediatric Medical University, Saint-Petersburg, Russia

The emergence of COVID-19 has posed challenges for healthcare professionals to quickly diagnose and provide timely medical care to patients. The basis for making managerial decisions on the tasks set is statistical accounting. [1] In this work, a comparative statistical analysis of the incidence of ARVI, new coronavirus infection COVID-19 and community-acquired pneumonia was carried out in an area with a population of 240,809 people. (184490 people - adults over 18 years old; 56319 - children from 0 to 18 years old). Medical assistance to the population was organized in accordance with the methodological recommendations of the Russian Ministry of Health. [2] To collect operational information, a form for recording the incidence of the adult and child population with acute respiratory viral infectious diseases, new coronavirus infection COVID-19 and community-acquired pneumonia was developed and implemented. When registering these diseases, the recommendations of the Ministry of Health of the Russian Federation (MHRF) were used. [3] The incidence rate and the number of those observed in the medical organization as of the current date during 2020 was recorded daily, starting from 04.20.2020 to 12.31.2020. It was found that the total number of people with ARVI, new coronavirus infection COVID-19 and community-acquired pneumonia observed in pediatric and adult clinics has two «waves» of rise. The minimum number of these diseases was registered in the 13th week (from 20.07.2020 to 27.07.2020)

for COVID-19 and ARVI, for community-acquired pneumonia - in the 17th week (from 17.08.2020 to 23.08.2020). In the structure of the incidence of COVID-19 in the first «wave», adult patients prevailed 93.3%, children - 6.7%. During the second «wave» of the rise in the incidence of COVID-19, the proportion of children doubled to 12.9%. The total incidence of COVID-19 in the first «wave» was recorded at 7.3 per 1000 people, in the second - 31.4 per 1000 people. The structure of the incidence of ARVI when comparing the two periods is different. In the first «wave» the number of adult patients prevailed - 70%, children - 30%. During the second «wave», ARVI was mainly registered in children - 71.7%. The total incidence of acute respiratory viral infections in the total adult and child population in the first «wave» was 29.4 per 1000 people, in the «second» - 154.7 per 1000 people. Community-acquired pneumonia for the entire period was registered in 2725 cases, of which in 4 children (0.14%), in 2721 adult patients (99.8%). The total incidence of community-acquired pneumonia in the total adult and child population in the first «wave» was 6.8 per 1000 people, in the «second» - 4.5 per 1000 people. The increased infectious morbidity required the involvement of additional medical personnel, transport, as well as the introduction of new organizational technologies for providing medical care to the population (mobile medical teams, dispensing drugs and pulse oximeters for providing medical care at home). The data of regular statistical observation became the basis for making operational management decisions for the organization of medical care for the population in the context of an epidemic rise in morbidity.

References:

1. Orel VI, Kim AV, et al. Scientific and practical vector of problems of primary medical and social care in a metropolis. Medicine and healthcare organization. 2018; 3 (2): 63–67.

2. Ministry of Health of the Russian Federation, Temporary guidelines «Prevention, diagnosis and treatment of new coronavirus infection (COVID-19), version 10». Access: https://base.garant.ru/74810808/. Date of treatment 01.25.2021

3. Letter of the Ministry of Health of Russia dated 08.04.2020 No. 13-2 / I / 2-4335 «On coding of coronavirus infection caused by COVID-19» Access: http://docs.cntd.ru/document/564672780 Date of treatment 01.25. 2021

Keywords: acute respiratory viral infection (ARVI), new coronavirus infection COVID-19, community-acquired pneumonia, morbidity, medical supervision.

The impact of unemployment on people's quality of life in Greece

Isidora Ovezikoglou[1], Sonia Malefaki[2]

[1]School of Sciences and Technology, Hellenic Open University (HOU), Greece,
[2]Department of Mechanical Engineering & Aeronautics, University of Patras, Rio Patras, Greece

Nowadays, unemployment is considered one of the most important macroeconomic problems that most developed countries are facing. Nevertheless, it remains a serious social problem as it generally affects the quality of people's life and their families. Quality of life is a multidimensional concept, which refers to various aspects of physical, mental and social well-being, as well as of the individual's life. Recently, the scientific community has shown a great interest on determining the indicators that best describe and analyze the quality of people's life. The main aim of the current study is to investigate whether unemployment affects the quality of people's life in Greece. This survey was based on the use of quality of life indicators selected from the European Union (EU) 8 + 1 quality of life framework and examined their relationship with the unemployment rate. Data on quality of life indicators and unemployment are available on an annual basis, covering the period 2003-2018 for the quality of life indicators and 1974-2018 for unemployment, published by Hellenic Statistical Authority (ELSTAT). Using multivariate statistical methods, such as principal component analysis (PCA), an overall index of quality of life was extracted and its relation with unemployment was studied. Finally, an attempt was made to identify the most appropriate prediction model of unemployment in Greece using the typical ARIMA models, which are considered extremely flexible tool for predicting the unemployment rate. The results of the analysis confirmed the predictive power of the models.

Keywords: unemployment, quality of life, quality of life indicators, principal components analysis, ARIMA models.

On the random signal detection problem in the quaternion domain

Antonia Oya
Department of Statistics and Operations Research, University of Jaén, Jaén, Spain

Problems involving quaternion random signals are commonly found in the field of statistical communication theory and they have recently attracted significant attention due to their multiple applications. They have shown to be useful, for example, in seismology, acoustics, electromagnetics, image processing, aerospace, etc. In all of these fields, the received signal is composed of a certain number of random components so it can be represented as a quaternion signal, which accounts for the correlated nature of its components. Similarly to the case of complex-valued random signals, the suitable statistical processing for quaternions signals requires the augmented statistics to be considered, i.e., the full second order statistical information carries along the three perpendicular quaternion involutions too. This approach is known as quaternion widely linear (WL) processing and has shown advantages over the traditional quaternion linear processing for a wide range of problems, which include adaptive filtering, blind extraction, and modeling, among others. In this paper, the problem of detecting a random signal corrupted by additive noise is reformulated in the quaternion domain and the WL processing is applied to derive a quaternion detector.
Keywords: Detection problem, quaternion random signals, widely linear processing.

Large Deviation for Conditional Volterra Processes

Barbara Pacchiarotti
Università degli Studi di Roma "Tor Vergata", Rome, Italy

In this work we investigate some problems of large deviations for continuous Volterra processes under the influence of model disturbances. More precisely, we study the behavior, in the near future after T; of a Volterra process driven by a Brownian motion in a case where the

Brownian motion is directly observable and where it is not directly observable (only a noisy version is observed or some linear functionals of the noisy version are observed). Some examples are discussed

Keywords: Large deviations, Volterra type Gaussian processes, conditional processes.

Validation of a Scale of Intention to Travel to a Tourist Destination

Taís Alexandre Antunes Paes, Rodrigo Ladeira, Neir Antunes Paes

Federal University of Aracaju, Aracaju, Brazil

In the context of the growth of tourist destinations, the travel market has been changing over the years, with rapid transformations and innovations, making it more competitive. Therefore, studying consumers means understanding their needs, interests, motivations and expectations. The objective of this paper was to develop a scale of intention to travel for leisure to a tourist destination, considering the customization formed by two constructs. The first one was called motivation proposed by Swarbrooke and Horner (2016) formed by the variables: cultural, status, emotional, personal, physical and personal development. The second one was called valu, which was proposed by Schwarz (1994) formed by the variables: power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity and security. In all questions for the proposed variables, we sought to address the customization of tourist attractions and verify their influence on decision making. To measure such variables, the phrase completion scale was used, which was submitted to the evaluation of eleven specialists. A questionnaire was prepared and submitted to a pre-test with 30 tourists, considering two big size cities in the northeastern region of Brazil (Aracaju and Salvador). To validate the scale, confirmatory factor analysis was used. After validating the scale, 215 people were surveyed in the city of Aracaju and 240 in Salvador, with an interest in traveling for leisure. The validation of the constructs or factors, based on all the reliability and validity indicators used, showed that the interviewees' assessment of the intention to go to tourist destinations was valid and reliable. For the value

construct, the ten proposed variables were validated and for the motivation construct, two variables were removed from the proposed model: cultural and personal development. At the end, it is possible to propose a validated and unprecedented questionnair! e in the area of consumer behavior and tourism.
Keywords: Marketing, Tourism, Destinatio.

Assessing the predictive ability of subjective survival probabilities

Apostolos Papachristos[1], Georgia Verropoulou[2]
[1,2]Department of Statistics and Insurance Science, University of Piraeus, Greece

Subjective survival probabilities vary based on socio-demographic and health factors and they reflect own views of future survival. The objective of this study is to investigate whether subjective survival probabilities can predict actual mortality. To achieve this objective we introduce the concept of 'Force of subjective mortality' and use it to estimate subjective survival probabilities related to a prediction time interval of two years. We then use a longitudinal dataset from the 6th and 7th Waves of the Survey of Health, Ageing and Retirement in Europe. For the statistical analysis we employ Binary Logistic regression and Cox regression models. Our results show that subjective survival probabilities contain supplementary information about own survival, in addition to socio-demographic, health and lifestyle factors, which are known to predict mortality. Moreover, the incorporation of 'Force of subjective mortality' in the analysis allows estimate subjective survival probabilities for different prediction intervals and therefore facilitates the use of data form consecutive SHARE Waves.
Keywords: Subjective survival probabilities, SHARE, self-rated health, force of mortality.

I-Binomial process of order k

Silvana Paralloi[1], Leda D. Minkova[2]
[1]Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria, [2]Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Bulgaria

In this paper we introduce a new compound binomial process with truncated geometric compounding distribution. The counting process is defined as a compound pure birth process. The equivalence of the definitions is given. The probability mass function, recursion formulas and some additional properties are derived. As application, we consider the discrete time risk model with the corresponding counting process, called I-Binomial risk model of order k.
Keywords: compound binomial process, birth process, I-Binomial risk model of order k.
Acknowledgements: The second author was totally supported by project DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

Component-based method for mixed data. Application in biology

Martin Paries[1], Evelyne Vigneau[1], Stephanie Bougeard[2]
[1]StatSC (USC) Statistique-Sensométrie-Chimiométrie Oniris Nantes, Rue de la Géraudière, Nantes, France, [2]Anses (National Agency for Food, Environmental and Occupational Health and Safety), Ploufragan, France

In modern data science, there is a growing need for multivariate statistical methods that deal with variables from various origin as well as various nature (i.e., numeric, nominal, ordinal). The main available methods can be split into three categories: (i) Linear methods derived from Principal Component Analysis (PCA) applied to pre-processed numeric and nominal variables, such as PCAmix method [Kiers, 1991] or Factorial Analysis for Mixed Data (FAMD) [Pagès, 2004]; (ii) Methods developed for nominal variables, e.g. Multiple Correspondence Analysis (MCA), with numeric and ordinal variables coded in pseudo-disjunctive forms [Escofier, 1979]; (iii) Optimal Scaling methods based on quantification

adapted to the nature of the variables using an Alternating Least-Square procedure [Young, 1981], e.g. PRINCALS [De Leeuw, 1980].
The aim of this work is to: (i) present these methods in a unified framework and (ii) compare them on the basis of a real dataset pertaining to the biological field.

References:
Kiers, H. A. (1991). Psychometrika, 56(2), 197-212.
Pagès, J. (2004). Revue Statistique Appliquee, 52, 93-111.
Escofier, B. (1979). Cahiers de l'Analyse des Données, 4(2), 137-146.
Young, F. W. (1981). Psychometrika, 46(4), 357-388.
De Leeuw, J., & Van Rijckevorsel, J. (1980). Data analysis and informatics, 2, 231-42.
Keywords: Mixed data, PCA, pre-processing, optimal scaling, quantification.

A Recursive Optimization Algorithm for Optimal Multiple Change-Point Detection and Inference

Christina Parpoula[1], Alex Karagrigoriou[2]
[1]Department of Psychology, Panteion University of Social and Political Sciences, Athens, Greece, [2]Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovasi, Samos, Greece

Change-point detection is the task of finding changes in the underlying model of a signal or time series. This paper addresses the problem of “a posteriori” optimal change-point detection and inference, performing a global segmentation. Under this framework, a recursive optimization algorithm is developed for selecting the optimal number of segments subject to the constraint of a minimum segment length, given a stream of process data. This procedure provides an empirical reasonable compromise for fine tuning these two parameters. Further, a multiple nonparametric change-point model along with a permutation approach to hypothesis testing is adopted for statistical inference. The method works as a mean shift detector allowing control the false alarm probability without

any knowledge about the specific distribution from which the observations are drawn. It therefore addresses a wide class of real-life contexts and problems where the identification of optimal level shifts in a time series is the main goal. Simulations and applications to real data are presented.

References:

Capizzi G and Masarotto G. Phase I distribution-free analysis of univariate data. J. Qual. Technol. 2013; 45:273-284.

Lavielle M and Moulines E. Least-squares Estimation of an Unknown Number of Shifts in a Time Series. J. Time Ser. Anal. 2000; 21:33-59.

Keywords: Change Point Detection, Dynamic Programming, Optimization, Segmentation, Signal, Time Series.

A multisite-multivariate AQI and the determination of new threshold values for health risk categories

Giuliana Passamani[1], Paola Masotti[2]

[1,2]Department of Economics and Management, University di Trento, Trento, Italy

Assessing air pollution and determining its impact on human health have been the focus of many studies. Various air quality indices (AQIs) have been proposed and analysed on observed data: given that pollution is a phenomenon described by several variables jointly, AQIs aim at combining diverse pollutants, characterized by different order of magnitude and unit of measure, in order to provide a simple and comparable indicator properly describing air quality. A further complication arises when we need to aggregate pollution data measured at different monitoring sites located over a certain geographical area. In the case of a three-dimensional array containing daily data on diverse pollutants at multiple sites, the aggregation procedure can be performed either by first reducing the space dimension, or by first managing the variety of the pollutants. When considering the reduction of data dimensions, principal component analysis (PCA) is a candidate statistical technique: three-way PCA can be applied symmetrically or asymmetrically, allowing an easier interpretation of the data structure. When we apply the technique on pollution data, we obtain a spatial synthetic aggregate AQI that should avoid the problems of ambiguity and eclipsicity (Swamee and Tyagi, 1999) in order to be reliable. Aggregation combining different pollutants at multiple sites leads to an AQI that will take on values measuring the

conjoint effect of the pollutants over a given area, and therefore higher than the values observed for the single pollutant. Official agencies indicate health risk threshold values for any single pollutant but not for the possible combinations of them, where values representing new thresholds have to be generated by the same technique used for aggregating the observed data. The aim of the present paper is the suggestion of a procedure for the determination of proper health risk categories for the overall aggregated AQI. A reliable AQI is of utmost importance for studying associations with respiratory and cardiovascular diseases, as well as with **Covid-19 spread.**

Keywords: Aggregate air quality index, Principal component analysis, Aggregate threshold values, Pollution health risk.

Hidden Markov model to analyze NEET and youth unemployment rates comparing EU countries over time

Fulvia Pennon[i1], Beata Bal-Domańska[2]

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy, [2]Department of Regional Economy, Wroclaw University of Economics and Business, Poland

European Union (EU) and governments are monitoring the labor market of young people since many years. We employ a Hidden Markov model to detected the dynamics of those who are Not in Education, Employment or Training (NEET) as well as the Youth Unemployment (YU) rates for the 28 EU countries over a period ranging from 2004 to 2018. The model relies on an unobservable Markov chain to account for the dependence between the observed rates at different time-occasions. The influence of key macroeconomic factors is accounted by considering a suitable parameterization to estimate their effects on the probability of a country to be in a certain state at the first time occasion and to move between states later on time. Maximum likelihood estimation of the model parameters is carried out through the Expectation-Maximization algorithm. Significant effects towards positive improvements in the examined context are estimated for the percentage of part-time contracts, the participation rate in education and training and the GDP growth. Through the decoded

states we show that Italy is the worst performing country over time since it is always allocated in the cluster having the highest NEET rates along with Bulgaria. Italy is also the worst country in terms of YU rates along with Greece and Spain.

Keywords: Expectation-maximization algorithm, classification, labor market, longitudinal continuous data, young people.

High Speed and Secured Network Connectivity for Higher Education Institution Using Software Define Networks

Lincoln S. Peter, Viranjay M. Srivastava
Department of Electronic Engineering, Howard College,
University of KwaZulu-Natal, Durban, South Africa

During Covid-19, there is a demand for high data throughput and connectivity from higher education institutions. This requires backbone and metro links to be upgraded with more capacity, which cater to these high data demands and connectivity. In addition, reliable and secured network is also a requirement against attacks from hackers. Cyber infrastructures need to be put in place to secure the network and prevent information from being accessed by unknown users and outsiders. Also, it has been observed that some network devices lose configurations with power failures around the areas. Due to this, several of the devices even lose routing and signaling information. This requires a technician to reconfigure the whole device manually. Also, throughout this process, there have been a number of challenges on the deployment of transmission network equipment. To solve these issues a model has been proposed, which is deploy on the Software Defined Network (SDN). This network will have three layers, namely: application, control, and data. Typically, the OSI model has seven layers, but some layers have been combined and reduced to three in this proposed model. This SDN will be user-friendly as it will be programmable to execute some of the tasks. It also saves bandwidth as it will reuse network resources. In the case where power (electricity) failed, and the networking device is rebooting, it will

automatically look up for configuration information on the SDN server and compare it with the one configured on the device. If this comparison is the same, then the device will work as normal. If the network device lost configurations, then the SDN server will automatically reload the device's configurations. The SDN will also do calculations of the bandwidth utilization for all the links or routes connected to it. This assists to see where network resources are used the most. If a particular route is congesting, the SDN will look for another alternative route where utilization is low so that it can load balance the traffic.

Keywords: Cyber system, Devices, Higher education, Network, Software defined radio.

Portfolio Optimization under Correlation Constraint

Traian Pirvu

Mathematics and Statistics, Faculty of Science, McMaster University, Ontario, Canada

We consider the problem of portfolio optimization with a correlation constraint. The framework is the multi-period stochastic financial market setting with one tradable stock, stochastic income, and a non-tradable index. The correlation constraint is imposed on the portfolio and the non-tradable index at some benchmark time horizon. The goal is to maximize a portofolio's expected exponential utility subject to the correlation constraint. Two types of optimal portfolio strategies are considered: the subgame perfect and the precommitment ones. We find analytical expressions for the constrained subgame perfect (CSGP) and the constrained precommitment (CPC) portfolio strategies. Both these portfolio strategies yield significantly lower risk when compared to the unconstrained setting, at the cost of a small utility loss. The performance of the CSGP and CPC portfolio strategies is similar.

This is joint work with Aditya Maheshwari.

Asymptotic Results for the Stochastic Inventory Model of Type (s, S) with Dependent Components

Aynura Poladova[1], Salih Tekin[1], Tahir Khaniyev[1]
[1]Department of Industrial Engineering, TOBB University of Economics and Technology, Ankara, Turkey

In this study, stochastic inventory model of type (s, S) is mathematically constructed and investigated under the assumption that the amount of demands and inter-arrival times are dependent to each other. This condition makes this study different from other studies in this field. This assumption is important for modeling a number of interesting problems arising in reliability, queuing, inventory, stock control theories, mathematical insurance, financial mathematics, mathematical biology and physics. In the study, it is supposed that the initial and control levels of the stock in a depot are equal to S and $s \geq 0$, respectively. Moreover, the amount of stock in the depot decreases by a random amount at random times, until the amounts of stock in the warehouse falls below a certain control level $s \geq 0$. When the amount of stock in the depot drops below the control level $s \geq 0$, we refill the stock immediately to initial level S. Thus, the first period is completed and the system demonstrate similar behavior afterward. The process expressing this model is called as a stochastic process with a dependent component (X(t)) in the literature. In this study, this stochastic process X(t) is constructed mathematically and stationary characteristics of the process are studied. Namely, the exact expression for the ergodic distribution of the process X(t) is derived. Moreover, an asymptotic expansion for the ergodic distribution of the process X(t) is obtained under the assumption that the dependency is linearly. In addition, the weak convergence theorem for the ergodic distribution of the process is proved. Moreover, the exact expressions and asymptotic expansions for the nth order moments ($n = 1,2,3, \ldots$) of ergodic distribution of the process X(t) are obtained, as $S-s \rightarrow \infty$.

Keywords: Stochastic Inventory Model of Type (s, S), Ergodic Distribution; Weak Convergence, Linear Dependence, Asymptotic Expansion.

Extreme value parameters estimation: an application to environmental data

Dora Prata Gomes[1], Helena Penalva[2], Sandra Nunes[3], M. Manuela Neves[4]

[1]Faculdade de Ciências e Tecnologia and CMA/FCT, Universidade Nova de Lisboa, Portugal, [2]Escola Superior de Ciências Empresariais - Instituto Politécnico de Setúbal and CEAUL, Universidade de Lisboa, Portugal, [3]Escola Superior de Ciências Empresariais - Instituto Politécnico de Setúbal and CMA/FCT, Universidade Nova de Lisboa, Portugal, [4]Instituto Superior de Agronomia, and CEAUL, Universidade de Lisboa, Portugal

Modeling the behavior of rare events is crucial in many areas of application such as, hydrology, insurance and environmental studies, where natural disasters can have strong impacts. There are several parameters that need to be estimated in many applications of extreme value theory (EVT). The extreme value index (EVI) is a parameter of primordial importance and the basis for the estimation of all other parameters of extreme events. Other challenges in extreme value analysis is modeling and estimating clusters of extreme values since they are linked with incidences and durations of catastrophic phenomena. Here, an important parameter comes into play, the extremal index (EI) that characterizes the degree of local dependence in the extremes of a stationary sequence. It needs to be adequately estimated, not only by itself but because its influence on other relevant parameters, such as high quantiles. After a brief overview of the developments on the estimation of parameters of extreme events we will use a data set in the environmental area to show how the EI estimation can affect the estimation of a high quantile.

Keywords: Extreme Value Index, Extremal Index, High Quantile, Environmental data.

Acknowledgments: This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications) and the project FCT-UIDB/00006/2020.

Prediction of coagulation activity in patients with paroxysmal atrial fibrillation

Krasimira Prodanova[1], Maria Negreva[2]

[1]Faculty of Applied Mathematics&Informatics, Technical University of Sofia, Sofia, Bulgaria, [2]Department of Cardiology, Medical University of Varna, Varna, Bulgaria

Paroxysmal atrial fibrillation (PAF) is the most commonly diagnosed arrhythmia and it affects up to 3% of the total population. By the end of 2060, it is expected to affect around 18 million people in Europe. One of the major clinical and socio-economic problems that PAF causes are thromboembolic events. High morbidity and embolic risk arouse the interest in the coagulation changes occurring in PAF. The creation of mathematical models to determine coagulation activity in PAF patients would be a significant clinical contribution. This will allow for quantification of the coagulation process and thrombotic potential of the disease. Evaluation of diagnostic tests is a matter of concern in modern medicine not only for confirming the presence of disease but also to rule out the disease in healthy subjects. In diagnostic test with dichotomous outcome (positive/negative test results), the conventional approach of diagnostic test evaluation uses sensitivity and specificity as measures of accuracy of test in comparison with gold standard status. The receiver operating characteristic (ROC) curve is shown to be a simple yet complete empirical description of this decision threshold effect, indicating all possible combinations of the relative frequencies of the various kinds of correct and incorrect decisions. The aim of this study is to quantify the effect of fibrinolisa factors to PAF manifestation. In the present work, using a dichotomous univariate logistic regression analyssis, estimates of the probabilities of PAF are derived. As predictors for PAF in the constructed logistic models the following parameters have been analyzed: Plasminogen level, t-PA level, PAI-1 activity, D-dimer, vitronectin and -antiplasmin acivity. Values $p<0,05$ were adopted for statistically significant. Statistical modeling of specific risk factors allows for potential identification of patients with PAF at risk and may help in formulation of future therapy strategies. Fibrinolisa factors to PAF manifestation, logistic regression models, receiver operating characteristic.

Acknowledgements:

The first author is totally supported by project DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

Sensitivity analysis and tail variability for the Wang's actuarial index

Georgios Psarrakos, Polyxeni Vliora
Department of Statistics and Insurance Science, University of Piraeus, Greece

The presentation refers to comparison of actuarial indices. Measuring and predicting risks (or losses) plays an important role in Actuarial Science. The survival function of a random variable is one of the most basic ways of calculating risks. The ranking of insurance risks with respect to their right tail is a challenging problem. To be more specific, in this paper, we extend the actuarial index introduced by Wang (1998) and propose its sensitivity index based on Leser's perturbation analysis on a proportional hazards model. We use tail variability measures by conditioning the risk for values greater than the Value-at-Risk (VaR), also we study in details how the VaR affects the actuarial and the sensitivity index. We provide characterization results for Pareto and exponential distributions, two cases where the actuarial and its sensitivity index are independent from VaR. We also obtain monotonicity results and bounds for them. The results are illustrated by numerical examples.
Keywords: Right-tail index; Tail variability measures; Proportional hazards model; Cumulative residual entropy; Mean residual lifetime; Quantiles; Value at Risk.

Mutual information rate, Kolmogorov-Sinai entropy and synchronization in cyclic networks with discontinuous local dynamics

J. Leonel Rocha[1], S. Carvalho[2]
[1]CEAUL. ADM, ISEL-Eng. Superior Institute of Lisbon, IPL, Lisboa, Portugal,
[2]CEAFEL. ADM, ISEL-Eng. Superior Institute of Lisbon, IPL, Lisboa, Portugal

The mathematical information theory studies the quantification, storage and communication of information, highlighting the quantities that are known as information measures, their properties and applications. In addition to its clear importance in the area of telecommunications, information theory has several applications in other scientific and technological areas such as: biology (computational neuroscience),

physics (quantum computing), chemistry (intercellular communication) and mathematics (statistical inference, cryptography, network theory and graph theory). The mutual information rate measures the amount of information produced by a network. This rate measures the flow of information between channels connected in a network. Another crucial measure in information theory is the Kolmogrov-Sinai entropy, which quantifies the uncertainty that defines information. These measures depend on several factors, namely the network topology and the local dynamics, and both are expressed in terms of the conditional Lyapunov exponents. It is also a well known fact that chaotic systems can be synchronized. Motivated by the theoretical and practical connection between the information measures and the synchronization phenomenon, our purpose in this work is to analyze the relations between the mutual information rate, the Kolmogorov-Sinai entropy and the synchronization in the cyclic networks of order N, since information theory and synchronization are directly related in a network. The discontinuous local dynamics considered at each node establishes the topological, metrical and chaotic complexity of the network that is being studied. A similar study has already been performed for complete dynamical networks, see [3], [4] and [5]. Discontinuous dynamical systems are recurrently found in physical systems and are used in several applications in various fields such as engineering, economic, biological and ecological models, among others. The study of discontinuous dynamics in synchronization phenomena has also attracted the attention of several researchers, see, for example, [1] and [2].

References:

1. V. Avrutin, L. Gardini, I. Sushko, F. Tramontana, Continuous and Discontinuous Piecewise-Smooth One-Dimensional Maps: Invariant, Sets and Bifurcation Structures, World Scientific, Singapore, 2019.

2. A. C. J. Luo, Discontinuous dynamical systems and synchronization, Eur. Phys. J. Spec. Top., 228 (6) (2019), 1383–1384.

3. J. L. Rocha, S. Carvalho, Information theory, synchronization and topological order in complete dynamical networks of discontinuous maps, Math. Comput. Simul, 182 (2021), 340-352.

4. J. L. Rocha, S. Carvalho, Information measures and synchronization in regular ring lattices with discontinuous local dynamics, in Proceedings of CHAOS 2020, 13th Chaotic Modeling and Simulation International Conference, ISAST, Ed. Skiadas, C.H., (2020), 641-653.

5. J. L. Rocha, S. Carvalho, Complete dynamical networks: synchronization, information transmission and topological order, Discontinuity, Nonlinearity, Complex to appear (2021).
Keywords: cyclic dynamical networks; discontinuous dynamics; synchronization; information theory; Lyapunov exponents.
Acknowledgments: Research funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through the projects UIDB/00006/2020 (CEAUL), UIDB/04721/2020 (CEAFEL) and ISEL.

On Reliability of a Double Redundant System under Full Repairment Scenario

Vladimir Rykov[1,2], Nika Ivanova[1,3], Emmanuel Nibasumba[1]
[1]Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia, [2]Kharkevich Institute for Information Transmision Problems of Russian Academy of Sciences (IITP RAS), Russia, [3]V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

For two probability distributions and in the paper [1] a modified Laplace-Stieltjes transform has been introduced as follows In terms of this transform the analytical expressions for the main reliability characteristics of a double redundant system with arbitrary distributed life- and repair times of system components under partial repairment scenario has been found. In this paper we prolong investigation of the same model under the full repair components scenario. The analytical expressions for the time-dependent and the steady state system probabilities as well as its reliability function are represented in the paper. Proposed approach and obtained analytical results allow also to investigate the sensitivity of the system reliability characteristics to the shape of the system components life- and repair time distributions. Results of the sensitivity analysis are illustrated by several numerical examples. The work has been supported by RFBR Grant No. 20-01-00575 (recipients Vladimir Rykov, and Nika Ivanova).

References:

[1] V. Rykov. On Reliability of a Double Redundant Renewable System. // Springer Nature Switzerland AG 2020. M. Gribaudo et al. (Eds.): ASMTA

2019, LNCS 12023, pp. 1–9, 2020. https://doi.org/10.1007/978-3-030-62885-7_3.
Keywords: Double redundant system, Reliability characteristics.

Limit Theorems for Branching Random Walks with Heavy Tails

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

We consider a continuous-time symmetric branching random walk on as $Z^d, d \geq 1$, as a system of particles that can move, give offspring, or die at the origin, called branching source. The process starts with one particle at the point $x \in Z^d$. The particle performs a random walk until it reaches the branching source. Let us describe the evolution of the particle at the source in a short time tending to zero: the particle can jump to another point of the lattice, can give a random number of offspring, die or nothing happens to the particle. Then each newborn particle continues a random walk and branching according to the same laws, independently of other particles. The random walk is assumed to be symmetric, irreducible, spatially homogeneous so that the transition intensity a(x,y) from point to $x \in Z^d$.point y∈Z^d can be expressed as a function of one argument a(y-x):=a(x,y). The heavy tails condition is imposed on the transition intensities: for |z|→∞ we have a(z)~H(z/|z|)/|z|^(d+α), where α∈(0,2) and H(.) is a positive symmetric bounded function on sphere Sd-1. Thus, the variance of random walk jumps becomes infinite. The generation of particles at the source is described by a Markov branching processes.
The theorem on the asymptotic behavior of the survival probability of the population of particles depending on the dimension of the lattice d and the parameter α is proved, a classification of the asymptotic behavior of the moments of the number of particles at each point of the lattice and on the entire lattice depending on d and the parameter α is obtained.
The research was supported by the Russian Foundation for the Basic Research (RFBR), project No. 20-01-00487.
Keywords: Branching random walk, Multidimensional lattices, Limit theorem, Intermittency.

Effect of SARS-CoV-2 virus on life expectancy

Tamara P. Sabgayda[1,2], Alexandr V. Zubko[1,2], Viktorya Semyonova[1,2]
[1]Federal Research Institute for Health Organization and Informatics of Ministry of Public Health of Russian Federation, Moscow, [2]Institute for Demographic Research – Branch of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow

According to preliminary estimates, in 2020, the loss of life expectancy in Russia was 1.57 years due to SARS-CoV-2 infection. Moscow mortality data show that in case of death from COVID-19, the average age was 68.7 ± 0.2 years for men and 74.8 ± 0.2 years for women. If the infection acts as a concomitant disease and death occurs from other causes, than the average age of death are differ from not infected ones. Statistically significant differences were found for circulatory diseases: age of death increased in men from 70.5 (70.42-70.56) to 71.9 (70.78-72.92) and decreased in women from 81.0 (80.94-81.03) up to 80.0 (79.06-80.91). The average age of death from diabetes mellitus in the presence of concomitant infection is 64.4 ± 3.0 years for men and 66.1 ± 4.1 years for women, and without infection, these indicators are 61.6 ± 0.9 and 68.3 ± 0.8 years, respectively. The average age at death from obesity is 55.6 ± 3.0 in men and 61.1 ± 2.6 in women with concomitant infection and 58.3 ± 0.8 and 65.4 ± 0.7, respectively, without infection. Perhaps, as the number of observations increases, these differences will become statistically significant. In deaths from circulatory diseases, the proportion of people aged 65 and over among men and women with concomitant SARS-CoV-2 is significantly higher than among those who are not infected. In death from diabetes and obesity, the proportion of older men is significantly higher, and the proportion of older women is less in the case of co-infection. Thus, the loss of life expectancy is due not only to death from COVID-19, but also to the effect of the virus on the development of pathology in a number of chronic diseases. This influence is not unambiguous and not the same for men and women.

Keywords: the average age of death, circulatory diseases, diabetes mellitus, obesity, concomitant infection.

Addressing models associated to commutative Jordan Algebras

Carla Santos[1], Cristina Dias[2], Célia Nunes[3], João Tiago Mexia[4]
[1,2,3,4]IPBeja and CMA-FCT-UNLisboa, Portugal

Models with orthogonal block structure, OBS, are linear mixed models whose covariance matrix is a linear combination of known pairwise orthogonal orthogonal projection matrices that add up to the identity matrix. Furthermore, when the covariance matrix commutes with the orthogonal projection matrix on the space spanned by the mean vector, the linear mixed models will have commutative orthogonal block structure, COBS. Imposing this commutativity condition, it is possible to move from optimal estimation for variance components of blocks and contrasts of treatments to least square estimators, that are the best linear unbiased estimators whatever the variance components. In this work, we highlight the possibility of presenting OBS, COBS and mixed models associated to commutative Jordan Algebra in a unified way.
Keywords: Commutativity, Jordan Algebra, Linear mixed models, OBS.

The impact of cesarean sections on neonatal mortality in rural-urban divisions in a region of Brazil

Carlos Sérgio Araújo dos Santos[1], Neir Antunes Paes[2]
[1,2]Federal University of Paraíba, Campus I - Lot. Cidade Universitária, João Pessoa, Brazil

Neonatal mortality represents, in the first seven days of life, about 70% of infant deaths, being the main responsible for maintaining the levels of the Infant Mortality Rate, both in Brazil and in the world, where these levels are below 18 deaths per 100 thousand live births. One of the important factors associated with neonatal infant mortality is cesarean section, which in Brazil reaches levels around 57%, one of the highest in the world. This study aimed to investigate the impact of cesarean sections on neonatal mortality in rural-urban divisions in the state of Paraíba, in the northeastern Brazil from 2009 to 2017. It was selected a set of maternal-Infant variables common to the microdata bases of the Ministry of Health's

Mortality and Birth Information System: Gender and Race/Child color, Age of mother, Mother's level of education, Number of children born alive, Number of children born dead, Type of pregnancy, Duration of pregnancy, Birth weight and Type of childbirth. After using the database linkage technique and the imputation of missing data of the variables, the Multilevel Logistic Model was applied, which considered two levels: each neonatal death (level 1) and rural-urban municipal characterization (level 2). Three modeling were carried out: Type of Childbirth variable was assumed to be a dependent variable; in the second and third models as independent. In the light of the microdata of 5,149 neonatal deaths that occurred in Paraíba from 2009 to 2017, the highest proportions of neonatal deaths in children born via cesarean section were found in urban municipalities. The modeling results revealed that the chance of neonatal death of a child born via cesarean section increases by 21% as the mother's level of education increases, reduces by 21% (protective effect) the greater the number of children born dead is, and increases 49% with the increase of multiple pregnancy. The area of residence was considered as one of the differen tiating determinants of neonatal mortality. The results reinforce the urgent need for more effective public and private policies to reduce the rate of elective cesarean sections to acceptable levels (10% - 15%) mainly in urban municipalities for a consequent reduction in neonatal mortality levels.
Keywords: Cesarean section, Neonatal Mortality, Regional Divisions, Multilevel Logistic Model.

Profile regression model for analysing antibody response to COVID-19 vaccines

Annalina Sarra[1], Adelia Evangelista[1], Tonio Di Battista[1], Damiana Pieragostino[2]

[1]Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University "G.d'Annunzio" of Chieti-Pescara, Italy, [2]Center for Advanced Studies and Technology (CAST), University "G.d'Annunzio" of Chieti-Pescara, Italy

Coranavirus disease, emerged in 2019 in Wuhan (China), is a serious respiratory syndrome that has resulted in widespread morbidity and mortality. More than one year into the Covid-19 pandemic around 140

million people have been infected with the virus and over 3 million are the confirmed deaths across more than 200 countries. The coronavirus health emergency has urged the scientific community internationally to control the spread of the virus and find answers in terms of therapies and development of vaccines. Many studies, focusing on the vaccine efficacy and effectiveness have been already carried out or are in progress. To identify the immune correlate of protection, in this paper we analyzed the antibody response in a sample of 300 volunteers who were subsequently allocated to receive the first dose of vaccines Pzfizer or AstraZeneca. Subjects involved in the analysis were re-called at different timeframes after the first injection of vaccine for re-determination of IgG levels. All participants were also surveyed regarding post-vaccination symptoms, age, gender, comorbidities and some eating habits. To handle with the complexity of covariate related effects on immunoglobulin antibodies, we followed an analytic strategy based on Bayesian Profile Regression. This modelling approach is expressly designed for clustering covariate profiles into groups and associate these groups, via a regression model, to the relevant outcome. For our working data, the revised methodology has allowed to identify clusters of subjects according to their antibody levels and disentangle the effects of explanatory variables on immune response to SARS-CoV-2.
Keywords: Bayesian Profile Regression, SARS-CoV-2, anti-S1 spike IgG levels, cluster profile.

Point and Interval Forecasts of Death Rates Using Neural Networks

Simon Schnürch[1,2], Ralf Korn[1,2]

[1]Department of Financial Mathematics, Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany, [2]Department of Mathematics, University of Kaiserslautern, Germany

The Lee-Carter model has become a benchmark in stochastic mortality modeling. However, its forecasting performance can be significantly improved upon by modern machine learning techniques. We propose a convolutional neural network architecture for mortality rate forecasting, empirically compare this model as well as other neural network models to the Lee-Carter model and find that considerably lower forecast errors are

achievable for many countries in the Human Mortality Database. We provide details on the errors, forecasts and global behavior of our model to make it more understandable and, thus, more trustworthy. As neural networks by default only yield point estimates, previous works applying them to mortality modeling have not investigated prediction uncertainty. We address this gap in the literature by implementing a bootstrapping-based technique and demonstrate that it yields highly reliable prediction intervals for our neural network model. Finally, we find that annuity values can strongly differ depending on the model, highlighting the important issue of model risk.
Keywords: Mortality forecasting, neural networks, convolutional neural networks, uncertainty quantification, prediction intervals, Lee-Carter model, mortality of multiple populations.

Exogenous risk factors as a reservoir of deaths from cardiovascular diseases

V.G. Semyonova[1,2], A.V. Zubko[2,1], A.E. Ivanova[1,2], T.P. Sabgayda[2,1], V.G. Zaporozhchenko[2]

[1]st - Institute for Demographic Research – Branch of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences (IDR FCTAS RAS), Russia, [2]nd - Federal Research Institute for Health Organization and Informatics of Ministry of Health of the Russian Federation (FRIHOI of MoH of the RF), Russia

Currently mortality from cardiovascular diseases (CVD) in Russia exceeds the Western European ones more than 3-fold. The highest gap is registered in people aged 30-44. With age, the gap decreases, reaching the minimum (2-fold) in people aged 75+, which is rather paradoxically as the risk of CVD death is the highest in older ages due to the natural aging. The study analyzes changes in age-specific CVD mortality in the context of risk factors. First, the gap between Russia and Europe in terms of CVD mortality is mainly accounted for by young ages. Second, the structure of CVD mortality is age-specific: in young ages it is due to "other heart diseases", primarily alcohol-related and unspecified cardiomyopathy, while with age, coronary heart disease (CHD) and cerebrovascular

diseases (CVD) prevail. Third, the age profile of CHD and CVD and alcohol-related and unspecified cardiomyopathy fundamentally differs with age profile patterns of these cardiomyopathies coinciding with ones for external death causes, in particular, alcohol and drug poisoning. Fourth, there is a strong likelihood of a drug-related component in deaths from cardiomyopathy, unspecified, as exemplified by Moscow. Therefore, CVD risk factors are clearly age-specific: in older ages, mortality is due to endogenous factors that can be characterized as the natural aging, while in young and middle ages it is due to exogenous factors, behavioral risk factors, primarily alcohol and drugs.

Keywords: Currently mortality, mortality from CVD, age-specific CVD mortality.

Predicting changes in depression levels following the European economic downturn of 2008
Main topic: Health state of a population: definition, modeling and estimates

Eleni Serafetinidou[1], Georgia Verropoulou[1]

[1]Department of Statistics and Insurance Science, University of Piraeus, Athens, Greece

The economic crisis occurring in Europe since 2008 has caused major changes to people's lives. Past studies found that mental health disorders have risen during periods of economic recession for both genders in Europe while others have supported that males are more vulnerable compared to females. The target of this study is to assess the depression imprint for a large sample of Europeans after the 2008 crisis. The sample studied in the analysis comes from the database of SHARE (Survey of Health, Aging and Retirement in Europe), a multidisciplinary longitudinal and cross national database including material regarding health, socioeconomic and demographic information of individuals aged 50 or higher, resident in several European countries. The selection of respondents included those participating both in wave 2, carried out in 2006 – 2007 and wave 6, completed in 2015, covering cross-national material in two time periods, just before and following the economic recession. For the purposes of the analysis multinomial logistic regression

models were applied for the total sample and separately by gender, using SPSS 20. Special attention is given to the concurrent factors being associated with the depression burden in older ages, covering different domains of life, before and after economic recession. Findings indicate that health predictors including mobility limitations, instrumental activities of daily living and long term illnesses had increased after 2015 for the total population of individuals, indicating worse health levels. Further, cognitive function had declined as well. Concerning factors leading to decreasing depression levels, the highest contribution is due to the reduction of limitations in instrumental activities of daily living and in mobility. Furthermore, better cognitive function and life satisfaction levels are related to a decline in depression levels. In general, factors associated with increasing depression levels are related to a worsening of the health indicators. Men are more vulnerable to mental disease due to an increase of limitations in instrumental activities of daily living while women due to an increase in long term illness. Worse levels in orientation in time and life satisfaction affect both sexes, increasing depression scores. For males other factors related to depression are facing more economic difficulties and living alone.

Keywords: gender, depression, economic downturn, Europe.

Acknowledgement: “This work has been completely supported by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI)”.

WITHOUT TITLE

Foad Shokrollahi

School of Technology and Innovations, Mathematics
Foad Shokrollahi, University of Vaasa, Vaasa, Finland

We consider so-called long-range dependent completely correlated mixed fractional Brownian motion (ccmfBm). This is a process that is driven by a mixture of Brownian motion (Bm) and a longrange dependent fractional Brownian motion (fBm) that is constructed from the Brownian motion via a Molchan–Golosov representation. We provide a transfer principle for the ccmfBm and use it to construct Cameron–Martin–Girsanov–Hitsuda theorem and prediction formulas. Finally we discuss the simulation of ccmfBm.

The WebRDS effectiveness as a sampling method among sexual minorities

Samuel Silva[1], Gilvan Guedes[1], Paula Miranda-Ribeiro[1], Kenya Noronha[2]

[1,2]Demography Dept., Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

This paper discusses the effectiveness of a Web-based respondent-driven sampling (WebRDS) among lesbians, gays, and bisexuals in Minas Gerais, Brazil. RDS is seen as an effective sampling method to access hidden and hard-to-reach populations. The main assumption underlying the method is its capacity to trace the connection through several waves to reduce the sample dependence on initial seeds. The procedure proves to be advantageous as the recruitment is made by peers, avoiding "masking" bias, and the number of coupons is limited, avoiding super-recruiters. The web-based version of the method (WebRDS) promised to be even more effective since it allows the referral chain to progress up to 20 times faster and increases the chances of response since the participation is anonymous and does not demand any personal contact. However, the implementation of the method in the real world is still challenging. Some studies indicate that the number of waves typically used in the RDS is insufficient to obtain the asymptotic unbiasedness because the preferential referral behavior by respondents leads to bias. To discuss the effectiveness of the WebRDS to sample the LGB population we use a WebRDS sample of 798 LGB individuals, aged 18 to 65 years, living in Minas Gerais, Brazil in 2019. The results show that the seven waves of recruitment were insufficient to obtain an unbiased sample, as the average path length was quite short: 1,6 for the full sample, 1,3 for lesbians, 1,1 for gays, and 1,8 for bisexuals. Furthermore, 43% of the sample were composed of gay referred by gay, among lesbians and bisexuals this behavior represented 2% and 3% respectively. Our results underscore the need for caution in the analysis using RDS since it reflects the social structure and the stigmas embedded in society, affecting the referral behavior and, consequently, the effectiveness of the inferences proposed by the method.

Keywords: WebRDS, sampling method, bias, hard-to-reach population.

Labor force ageing in the Czech Republic: the role of education and economic industry

Martina Simkova[1], Jaroslav Sixta[1]
[1]Prague University of Economics and Business, Prague, Czech Republic

It seems that the economic activity of both men and women workers is decreasing with their age. However, the issue is more complex since the rate of economic activity of older workers is very different in each industry. In addition, it depends on the level of education. We assume that workers with higher educational level will have a higher rate of economic activity than workers with lower educational level. This contribution deals with the analysis of economic activity of older workers by sex, level of education in individual industries of the Czech economy. Our estimates are based on detailed data from both the Labor Force Survey and Statistics on Income and Living Conditions for the last few years. We show that economic activity differs according to educational level and it is specific for each industry of the economy. As the Czech population is ageing, the government is proposing to raise statutory retirement age. We show that the setting of the retirement age should take place at the level of these analyzed factors, not uniformly.
Keywords: ageing population, educational level, economic activity rates, industry.

Financial literacy and longevity awareness: a preliminary investigation

Rosaria Simone[1], Mariarosaria Coppola[1]
[1]Department of Political Sciences, Federico II University, Naples, Italy

Many survey studies world-wide are tailored to assess financial literacy and its determinants. On these basis, several scores have been proposed to quantify respondents' degree of knowledge about financial issues and the riskiness of their behaviour. By putting altogether findings about

gender gap in life expectancy and in financial knowledge, it emerges that scarse literacy and planning affect mainly women's life in view of the circumstance that they face a relatively weaker wealth situation through life and are more exposed to longevity risk. The contribution is framed within a project aimed at the investigation of the assocation between longevity risk awareness and financial literacy. The starting point is the understanding of self-evaluation of survival probabilities at older ages and of financial knowledge by means of an adequate methodological setting to model both perception and uncertainty of rating survey data. Then, in the long run, the idea put forth in our contribution is to establish a 'longevity literacy' index in the wake of existing measures of financial literacy and well-being. Indeed, individuals' awareness of the phenomenon could enhance the applicability of flexible retirement schemes in order to cope with low fertility, decaying mortality rates and late entereing in the job markets.

Keywords: Financial literacy; Longevity risk; survey data.

On two Algorithms for Clustering Interval Data with the Weighted Generalized Affinity Coefficient

Áurea Sousa[1], Leonor Bacelar-Nicolau[2], Osvaldo Silva[3], Helena Bacelar-Nicolau[4]

[1]Universidade dos Açores and CEEAplA, Ponta Delgada, Portugal, [2]Universidade de Lisboa, Faculdade de Medicina, Institute of Preventive Medicine & Public Health and ISAMB/FMUL, Lisboa, Portugal, [3]Universidade dos Açores and CICS.NOVA.UAcores, Ponta Delgada, Portugal, [4]Universidade de Lisboa, Faculdade de Psicologia and Instituto de Saúde Ambiental da Faculdade de Medicina (ISAMB/FMUL), Lisboa, Portugal

The weighted generalized affinity coefficient appears to be an appropriate resemblance measure between elements – either statistical data units or variables - of complex datasets, sometimes issued from large databases. In this work, we apply and compare two different algorithms to compute the weighted generalized affinity coefficient when cluster analysis is carried out over a dataset described by interval valued variables, that is, whose values are intervals of the real axis.

Two well-known case-studies issued from the multivariate data analysis literature, concerning applications in the environmental and management fields, are used, to illustrate and discuss both procedures.

References:

Á. Sousa, F.C. Nicolau, H. Bacelar-Nicolau, and O. Silva, "Weighted Generalised Affinity Coefficient in Cluster Analysis of Complex Data of the Interval Type", Biometrical Letters, vol. 47, no. 1, pp. 45-56, 2010.

H. Bacelar-Nicolau, F.C. Nicolau, Á. Sousa, and L. Bacelar-Nicolau, "Clustering of Variables with a Three-Way Approach for Health Sciences", Testing, Psychometrics, Methodology in Applied Psychology (TPM), vol. 21, no. 4, pp. 435-447, 2014, doi: 10.4473/TPM21.4.5

H. Bacelar-Nicolau, F.C. Nicolau, Á. Sousa, and L. Bacelar-Nicolau, "Measuring Similarity of Complex and Heterogeneous Data in Clustering of Large Data Sets", Biocybernetics and Biomedical Engineering, vol. 29, no. 2, 2009, pp. 9-18.

Keywords: Cluster Analysis, Affinity Coefficient, Hierarchical Clustering Models, VL Methodology, Validation Measures.

4C Assessment Model during 4IR: Implications and Impact in the Environmental Science and Engineering

Surabhi Srivastava, Cristina Trois

Department of Civil Engineering, Howard College Campus, University of KwaZulu Natal, South Africa

In the era of Industrial Revolution 4.0, education will have an impact on the assessment process and its application according to the industrial prerequisites and requirements. Education during the 4IR will shift the major part of the learning responsibilities from the teachers to the technology. Therefore, the assessment process by teachers should support strengthening the 4IR as well as its impact on the environment. A 4C assessment model is proposed with an example for the assessment of environmental science and engineering graduates as their knowledge will be applied in assessing the 4IR impacts on the environment. Critical thinking on the environmental issues during the fourth industrial revolution

would be analysed by critical thinking with artificial intelligence that will make wise solutions towards the problem. The logical mind maps would be easier to analyse and the practical implementation would be instant. Communication skills in the 21st century the assessment of environmental engineering or science students must be assessed on the communication skills as they are prospective future leaders in the respective field. Collaborations with respective authorities a part of the assessment must be on making possible collaborations by the students should be analysed. Studies in environmental science and technology in the future will need massive data to prove the existence of future environmental issues and only statistical data will create facts in the field.
Creativity for community and environment as the study module or course must be assessed by their effect on the environment. This 4C model for assessment will respond to the need of students, putting greater emphasis on certain topics, repeating things that students have not mastered yet, and generally facilitating students to adapt in the future workplace. As the education assessment 4 C model will improve with analysing student's capability it will fit with the disabled students too. This future assessment process will offer great opportunities for remote, self-paced learning.
Keywords: Pollution, Climate change, 4IR and environment.

A quantification study on precarity in Greece with evidence from the EU-LFS

Glykeria Stamatopoulou[1], Dimitris Parsanoglou[2], Maria Symeonaki[3]
[1,2,3]Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece

Numerous efforts have been made in the past to capture the multifaced phenomenon of employment precarity. However only a very short body of literature provides quantitative methods that measure the degree and forms of precarity and a way to perform cross-national comparisons. In this paper we provide a composite indicator that captures the degree of precarity employment with raw data drawn from the European Labour Force Survey (EU-LFS), that combines indicators such as unemployment, seasonal and/or part-time workers, temporary contracts, hours worked

and self-employment. The methodology is performed for the case of Greece, with the latest at the time available data, i.e. for the year 2019, however it could be adopted for other EU member states with practically no amendments. The results reveal an important degree of employment precarity in Greece for the year 2019.

Keywords: precarity, job insecurity, EU-LFS, composite indicator, Greece.

Intergenerational class mobility in Greece with evidence from EUSILC [Poster]

Glykeria Stamatopoulou[1], Maria Symeonaki[2], Catherine Michalopoulou[3]

[1,2,3]Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece

In this work, we study the intergenerational social mobility in Greece, in order to provide up-to-date evidence on the changes in the mobility patterns for both men and women aged between 30-60 years. The main purpose is to examine the relationship between the socio-economic status of parents and their children and the evolution of the mobility patterns between different birth cohorts. The role of education is also explored in shaping the mobility patterns. For the analysis, we draw data on both fathers' and individuals' social outcomes from different databases, and we use the European Socio-economic Classification (ESeC) and the International Standard Classification of Education (ISCED), in order to measure the social class of origins and destination. Applying the Markov transition probability theory, and a range of measures and models, this work focuses on the magnitude and the direction of the movements that take place in the Greek labour market, as well as the level of social fluidity. Three-way Markov mobility tables are presented, where the transition probabilities between the classes of origins and destination are calculated for different cohorts. A range of absolute and relative mobility rates, as well as distance measures are also presented. Finally, going one step beyond the existing indices in the literature we propose a new index of measuring the intergenerational mobility, that weights the transitions

according to the distances between parental and individuals' social class and a comparison to existing indices is presented.

Keywords: Intergenerational mobility, social class, cohort analysis, Markov system, distances, mobility indices.

REDACS: Regional emergency driven adaptive cluster sampling or effective COVID-19 prevalence

Milan Stehlík[1,2]

1Department of Applied Statistics & Linz Institute of Technology, Johannes Kepler University in Linz, Austria, [2]Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

As COVID-19 is spreading, national agencies need to monitor and track several metrics. Since we do not have perfect testing programs on the hand, one needs to develop an advanced sampling strategies for prevalence study. The recent importance of COVID-19 mitigation strategies motivates necessity of scalable, interpretable and precise methodology, which has materialized as REDACS. In this talk we will discuss its feasibility of REDACS implementations. We introduce REDACS: "Regional emergency driven adaptive cluster sampling" for effective COVID-19 prevalence and justify its usage as COVID-19 mitigation strategy. We show its advantages over classical massive individual testing sampling plans. We also point out how regional and spatial heterogeneity underlines proper sampling. Fundamental importance of adaptive control parameters from emergency health stations and medical frontline is outlined. Since the Northern hemisphere entered Autumn and Winter season, practical illustration from spatial heterogeneity of Chile (Southern hemisphere, which already experienced COVID-19 winter outbreak peak) is underlying the importance of proper regional heterogeneity of sampling plan. We explain the regional heterogeneity by microbiological backgrounds and link it to behavior of Lyapunov exponents. We also discuss screening by antigen tests from the perspective of "on the fly" biomarker validation, i.e. during the screening.

References:

M. Stehlík, J. Kiselak, A. Dinamarca, Y. Li, and Y. Ying, On COVID19 outbreaks predictions: issues on stability, parameter sensitivity and precision, Stochastic Analysis and Applications DOI: 10.1080/07362994.2020.1802291
M. Stehlík, J. Kiselák, A. Dinamarca, E. Alvarado, F. Plaza, F.A. Medina, S. Stehlíková, J. Marek, B. Venegas, A. Gajdos, Y. Li, S. Katuscak, A. Brazinová, Y. Lu, REDACS: Regional emergency driven adaptive cluster sampling for effective COVID-19 prevalence, submitted.
Keywords: COVID-19, prevalence, testing, REDACS.

Alcohol Policy in the Czech Republic

Kornélia Svačinová[1], Markéta Pechholdová[2], Jana Vrabcová[3]
[1,2]Department of Demography, Prague University of Economics and Business, Czechia, [3]Department of Statistics and Probability, Prague University of Economics and Business, Czechia

Background: Czechia is one of the nations with the highest alcohol consumption in the EU and, conversely, low health literacy. An average Czech consumes 286 beers and 6.8 liters of pure alcohol per person per year. This is a serious problem with many health risks, e.g. occurrence of different types of cancer. Positive effect is caused e.g. by penalties in the case of excessive alcohol consumption or sale of alcohol to adolescents, increase in the price of alcohol, a ban on advertising and repression. Examples are Scandinavian countries, where there is a higher excise duty on alcohol and the consumer has limited options for buying alcohol. There is also a lack of education in Czechia, which would acquaint the public with the fact that alcohol, despite its easy availability, is a drug and its negative effects do not lag behind marijuana, tobacco and other addictive substances. Rather the opposite. Aims: This paper reviews documented action plans and regulations of the alcohol policy (and its gap) in Czechia (compared to other countries). Health consequences (diseases, mortality, morbidity) and calculations of selected indicators are conducted.
Keywords: Alcohol Policy, Alcohol Consumption, Czechia.

Capturing school-to-work transitions using data from the first European Graduate Survey

Maria Symeonaki[1], Glykeria Stamatopoulou[2], Dimitris Parsanoglou[3]
[1,2,3]Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece

The transition from education to work represents a critical stage in the lives of youth individuals. Over the past decades, transition into the working life has increasingly come to be seen as more varied and less standardised. The reasons for more varied transitions patterns are certainly multi-faceted. The present paper serves both descriptive and analytic aims. From a descriptive perspective we focus primarily on addressing the question of whether, how and to what extent transitions differ between European countries. Alongside, attempts to identify and explain the mechanisms behind different transitions patterns are explored. The analysis is cross-national with data drawn from the Eurograduate Pilot Survey, a one-off pilot survey of recent graduates in 8 European countries.
Keywords: school-to-work transition, employment patterns, youth unemployment, skills mismatch, Eurograduate Pilot Survey.

Pricing American Options Using a Semi-Markov Decision Process

Kouki Takada[1], Marko Dimitrov[2], Lu Jin1, Ying Ni[2]
[1]Department of Informatics, University of Electro-communications, Tokyo, Japan,
[2]Division of Mathematics and Physics, Mälardalen University, Västerås, Sweden

In this research, we consider the problem of American option pricing when the price dynamics of the underlying risky asset are governed by a varying economic situation. We assume that the situation of economic transits based on a semi-Markov process, and the pricing procedure is formulated using a semi-Markov decision process. The decision maker decides whether to early exercise or hold the option based on the information on both asset price and economic situation. The optimal strategies are investigated and some properties on the activity regions are discussed based on simulation results.
Keywords: Decision-making, Optimal policy, American option, Threshold type policy, Totally Positive of Order 2.

Consistent and asymptotically normal RGCCA estimators for linear structural equations with latent variables (RGCCAc)

Michel Tenenhaus[1], Arthur Tenenhaus[2], Theo Dijkstra[3]
[1]HEC Paris, France, [2]University of Paris-Saclay, CentraleSupelec, Laboratory of Signal and Systems, [3]University of Groningen, Netherlands

We show in this work how to use Regularized Generalized Canonical Correlation Analysis (RGCCA) in structural equation modeling for obtaining consistent and asymptotic normal estimators of the parameters. This new approach, named RGCCAc, is evaluated and compared to PLS-PM, PLSc, and SEM-ML on a simulation and on a data set referring to the European Customer Satisfaction Index (ECSI).
Keywords: Regularized Generalized Canonical Correlation Analysis, Structral Equation Modelling, Consistent Partial Least Squares.

Predicting Risk of Gestational Diabetes Mellitus through Nearest Neighbour Classification

Louisa Testa[1], Mark Anthony Caruana[1], Maria Kontorinaki[1], Charles Savona-Ventura[2]
[1]Department of Statistics and Operation Research, University of Malta, Tal-Qroqq Campus, Malta, [2]Department of Obstetrics and Gynaecology, University of Malta, Medical School, Malta

Gestational Diabetes Mellitus (GDM) might arise as a complication of pregnancy and can adversely affect both mother and child. Diagnosis of this condition is carried out through screening coupled with an oral glucose test. This procedure is costly and time-consuming. Therefore, it would be desirable if a clinical risk assessment method could filter out any individuals who are not at risk of acquiring this disease. This problem can be tackled as a binary classification problem. In this study, our aim is to compare and contrast the results obtained through logistic regression with the results obtained through the application of three well-known nonparametric instance-based learning classification techniques; namely,

k-nearest neighbours, fixed-radius nearest neighbours and kernel nearest neighbours. These three techniques were selected due to their relative simplicity, applicability to the problem being studied, lack of assumptions and nice theoretical properties. The dataset contains information related to 1368 mothers and their new-born babies across 11 Mediterranean countries. More specifically, the dataset consisted of 71 explanatory variables. However, following a variable selection technique, 30 variables were used. These included age, pre-existing hypertension, menstrual cycle regularity, history of diabetes in family and more. Since the data exhibited a class imbalance in favour of mothers that were not affected by GDM, SMOTE-NC was implemented prior to the application of the classification techniques to balance out the data. The performance of the above-mentioned techniques was evaluated using various performance measures (e.g., ROC curve, precision, accuracy, etc.), while cross validation techniques were used for the hyper-parameters' selection. The results revealed that nearest neighbour methods outperformed the logistic regression method that was previously applied in other studies.

Keywords: GDM, nonparametric techniques, k-nearest neighbours, fixed-radius nearest neighbours, kernel classifiers.

Robust Model Selection Criteria

Aida Toma[1,2], Alex Karagrigoriou[3], Paschalini Trentou[3]

[1]Department of Applied Mathematics, Bucharest University of Economic Studies, Bucharest, Romania, [2]"Gh. Mihoc – C. Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romanian Academy, Bucharest, Romania,

[3]Department of Statistics and Actuarial-Financial Mathematics, Lab of Statistics and Data Analysis, University of the Aegean, Karlovasi, Samos, Greece

We present a new class of robust model selection criteria. These criteria are defined by estimators of the expected overall discrepancy, using pseudodistances and the minimum pseudodistance principle. Theoretical properties of these criteria are presented, namely asymptotic unbiasedness, robustness, consistency, as well as the limit laws. The case of the linear regression models is studied and a specific pseudodistance based criterion is proposed. Monte Carlo simulations and applications for real data are presented in order to exemplify the performance of the new methodology. These examples show that the new selection criterion for regression models is a good competitor of some well

known criteria, and may have superior performance especially in the case of small and contaminated samples.

Keywords: Model Selection, Minimum Pseudodistance Estimation, Robustness.

Socio-economic factors and Infant Mortality Rate: Romania

Iuliana Toropoc
Universitatea Bucuresti, Academia de Studii Economice, Tehnici Actuariale, Repumlic of Romania

The association of socio-economic factors and Infant Mortality Rate (IMR) is well established. Romania is a developed country characterized by a high IMR. There are no studies to date on the relationship between socio-economic factors and IMR with Romanian population data. An understanding of the influence of socio-economic factors on IMR is beneficial, as the IMR is considered a valid measure of population health. A number of socio-economic indicators are selected from the database of the National Institute of Statistics. Those found to have significant associations with IMR are introduced in a regression model with IMR as independent variable. Our variables were: Net Average Income, Number of Doctors, Marriage Rate, Dropout Rate, Unemployment Rate. For Net Average Income, we used the Net Average Monthly Income which we converted into dollars and adjusted for inflation. For Unemployment Rate, we calculated our own crude unemployment rate, by dividing the number of yearly registered unemployed to the number of yearly registered employed. The data covered the time period 1990-2019, all counties, regions and microregions. Negative associations were found between IMR and Net Average Income, Unemployment Rate, Dropout Rate and Number of Doctors. Positive associations were found between IMR and Number of Doctors and Marriage Rate. For all data: At County level, Net Average Income was significant in 90% of the counties, while Marriage Rate in 33% of the counties. The effect of Number of Doctors was positive and significant for 9 of the 42 counties, while the effects of Number of Emigrants and Unemployment Rate were significant only for 6 and 3 of the 42 counties, respectively. The highest number of significant Independent Variables per county was 3, with the majority of the counties

featuring either 1 or 2 significant variables. Adjusted R-Squared values of .66-.93 indicated a good fit. At Region level, Net Average Income was significant for all regions and Marriage Rate for 90% of the regions. One region featured all 5 variables as significant. Adjusted R-Squared values between 0.84 and 0.67 indicated a good model fit. At Macroregion level, Net Average Income and Marriage Rate were significant for all the 4 regions, with MR2 being explained by all the 5 variables. Adjusted R-Squared values between 0.7! 3 and 0.81 indicated a good fit. The model was less successful when implemented at year level, this being confirmed by the low R-Squared values (0.41-0.15). Only 11 of the 30 years featured a significant variable, the most frequent variable being Unemployment Rate. For data featuring Dropout Rate: The Dropout Rate was significant in only 4 of the 42 counties, with one county featuring all variables as significant and an Adjusted R-Squared of .99. At Year level the Dropout Rate was significant only for year 2016. At Macroregion level, the Dropout Rate was significant only for MR4.
These results suggest that socio-economic factors have an effect on IMR and that targeted measures addressing these factors might result in a lowering of the MRI rate.
Keywords: Demographics, statistical models, infant mortality rate, socio-economic factors, Romania**.**

A faster No-U-Turn sampler for Hamiltonian Monte-Carlo

Samis Trevezas[1], Dimitra Blani[1], Andreas Markoulidakis2, Paul-Henry Cournède[3]

[1]Department of Mathematics, University of Athens, Athens, Greece, [2]Psychological Medicine & Clinical Neurosciences, School of Medicine, Cardiff University, Wales, UK, [3]MICS laboratory, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

Hamiltonian Monte Carlo (HMC) is a special type of Markov Chain Monte Carlo (MCMC) algorithm, which utilizes the Hamiltonian Dynamics to explore more effectively the support of the target distribution. The main drawback of this algorithm is the need to tune several parameters which are crucial for the performance of the algorithm and in particular, the step size ε and the desired number of leapfrog steps L. The No-U-Turn Sampler

(NUTS), proposed by Hoffman and Gelman (2004) is an algorithm which extends the idea of HMC, by introducing a slice variable, while tuning the crucial parameters of the latter uniquely. However, its performance is much slower, sometimes very poor and its basic version lacks reversibility. In this paper, we propose a Reversible edition of the No-U-Turn Sampler (R-NUTS), which avoids the use of a slice variable, and leads to a faster algorithm with simpler structure. Some simulated examples, including parameter estimation in a plant growth model, are also given to illustrate the performance of the algorithm.

Keywords: Hamiltonian Monte Carlo, No-U-Turn Sampler, reversible Markov chains, Markov Chain Monte Carlo.

Compositional data analysis with the R package Compositional

Michail Tsagris

Department of Economics, University of Crete, Gallos Campus, Rethymnon, Crete, Greece

Compositional data have gain popularity over the last years in many fields, such as geology, hydrochemistry, bioinformatics, political sciences, bioinformatics and economics. The subsequent increasing necessity for new methods and models for analyzing such data led to a growth of R packages that perform such analyses. Currently there are more than 20 R packages devoted to compositional data analysis, most of which include limited number of availabilities or are directed towards a single analysis.

On the contrary, the Compositional package offers various methods and techniques for analyzing and visualizing compositional data. The goal of this package is to serve as a channel for promoting compositional data in general. Effectively, the package embraces methods suggested by the its authors and methods suggested by other researchers working in the field. In this presentation I will first briefly overview the package and then focus on regression models for compositional data. Regression models for either compositional responses and predictors or their combination, delineating the models' supremacies and drawbacks.

Keywords: Compositional data, regression models, R.

Estimating mortality and reproductivity measures for the population of Greece at prefecture level 1981-2011: patterns and implications

Cleon Tsimbos[1], Georgia Verropoulou[2]
[1,2]Department of Statistics & Insurance Science
University of Piraeus, Greece

Aim In this paper we estimate for the first-time measures of fertility and reproductivity of the population of Greece at prefecture level for the period 1981-2011, we identify spatial patterns and discuss demographic implications.
Data. We use vital registration data on deaths by sex, age and place of residence of the deceased and the corresponding population counts based on the 1981, 1991, 2001 and 2011 censuses. The analysis is performed at prefecture level for each gender separately. To eliminate annual random fluctuations which may affect the mortality schedule, 3-year averages of the number of deaths occurring around the corresponding census dates are considered.
Method. Combining vital statistics with population data we calculate age-sex specific mortality rates for the periods 1980-1982, 1990-1992, 2000-2002 and 2010-2012. The observed mortality rates are graduated applying non-linear moving averages. The smoothed mortality rates are then used to construct regional abridged life tables based on the Chiang methodology. Subsequently, we apply the conventional Lotka theory to estimate regional Stable Population models.
Results. In 1981, 32 out of the 51 regions of the country exhibited TFR > 2.1 and r >0. However, since 1991, no administrative region reached population replacement levels. Between 1981 and 2011 the largest fertility decreases (more than one child or 41%) were observed in Kozani, Samos, Achaia and Rethymno. Regional differentials in fertility and reproduction are diminishing over time. Between 1981 and 2011 the national mean length of generation increased by 4.5 years (from 26.2 to 30.7 years); the largest increases were observed in Kastoria, Ioannina and Serres (6.5, 5.9, 5.7 years, respectively).
Conclusions. For the past two decades (1991-2011) all regions of the country experience rates below population replacement levels; this combined with the considerable increase in the mean length of the

generation, shape a very gloomy demographic future: under closed population conditions the population size will inevitably diminish while population ageing will become irreversible at national as well as regional level.

A Bi-level Model for Optimal Capacity Investment and Subsidy Design under Risk Aversion and Uncertainty

Maria Tsiodra[1], Michail Chronopoulos[1,2]

[1]City, University of London, Cass Business School, London, UK, [2]Norwegian School of Economics, Department of Business & Management Science, Bergen, Norway

Meeting ambitious sustainability targets motivated by climate change concerns requires the structural transformation of many industries, and, in turn, the careful alignment of firm- and Government-level policymaking. Indeed, while private firms rely on Government support to achieve timely the necessary green investment intensity, Governments rely on private firms to tackle financial constraints and technology transfer. Further complicating this interaction is that green energy technologies entail risk that cannot be diversified, and, thus, private firms may act more cautiously than Governments expect due to risk aversion. Therefore, we develop a bi-level, utility-based real options framework in order to analyse how a Government's policy on subsidy design interacts with a firm's investment decision under economic uncertainty and risk aversion. Our bi-level framework involves three main steps: i. We derive the firm's optimal investment policy, in terms of investment timing and project scale, assuming an exogenously defined subsidy; ii. Conditional on the firm's optimal investment policy, we solve the Government's optimisation objective to obtain the optimal subsidy level; iii. We derive the firm's optimal investment policy conditional on the optimal subsidy level.

Contrary to existing literature, results indicate that greater risk aversion lowers the amount of installed capacity yet postpones investment. Also, although greater economic uncertainty raises the optimal subsidy under risk neutrality, the impact of uncertainty is reversed under high levels of risk aversion.

Keywords: investment analysis, real options, subsidy design, risk aversion.

Comparison of Positivity in Two Epidemic Peaks of Covid-19 in Colombia with FDA

Cristhian Leonardo Urbano Leon[1], Manuel Escabias[2]
[1]Department of Mathematics of the Universidad del Cauca Sector Tulcán, CR 2A Nº 3N-111, Popayan, Colombia, [2]Department of Statistics and Operational Research. University of Granada, Spain

In the context of the global emergency due to Covid-19, the Colombian government has chosen the rate of positivity as one of the most important variables in making early decisions related to the management of the disease. The rate of positivity is the daily percentage of positive Covid-19 tests for the total of tests processed, its trend is used to help determine the presence of a peak epidemic outbreak. On the other hand, functional data analysis (FDA) is a statistical methodology whose main characteristic is that it focuses on the construction, treatment, and statistical analysis to continuous functions that come from variables with a continuous domain but which are discrete measures. Since the rate of positivity is a continuous variable with variations over time, in this work, the Covid-19 positivity curves are constructed for different regions of Colombia and statistically compared under FDA methodology taking as a reference the two peaks of an outbreak that occurred in Colombia between March 2020 and February 2021.
Keywords: Covid-19, FDA, Rates of Positivity.

Investigation of Fractional Brownian Fields

Neringa Urbonaite[1], Leonidas Sakalauskas[2]
[1,2]Vilnius University, Institute of Data Science and Digital Technologies, Vilnius, Lithuania

Much of the work using the fractional Brownian vector field (fBvf) model has been performed when the data are one-dimensional. The aim is to apply the fBvf model to vector data and to compare methods for estimating the Hurst parameter. A distance matrix and a new covariance function were introduced for the fBvf model. A Monte Carlo statistical experiment

was performed to analyze the model. An algorithm for fBvf generation and methods for estimating the Hurst parameter were developed. Two methods are chosen to estimate the Hurst parameter: maximum likelihood and the variogram. The maximum likelihood method, which is also used to make a forecast, is less likely to address this problem due to its complexity. The variogram method is widely used in geostatistics. A comparison of these methods in estimating the Hurst parameter allows us to conclude that the fBvf model is constructed correctly. This report provides a new definition of the multidimensional case of fBvf, with a new covariance function for which the starting point of the coordinates is not important, and justifies that the maximum likelihood method works better than the variogram. The practical application of the maximum likelihood method is performed to real temperature and precipitation data by investigating their dependences. These are steps that make it possible to simplify the modeling of multiannual data, at the same time, to analyze complex data more deeply and to predict their anomalies.
Keywords: Fractional Brownian vector field, Hurst parameter, maximum likelihood method, variogram.

Development of a New Bibliometric Indicator. Application in Dentistry

Pilar Valderrama[1], Pilar Baca[1], Mariano J. Valderrama[1]
[1]Department of Statistics, University of Granada, Faculty of Pharmacy, Campus Cartuja, Granada, Spain

The Journal Impact Factor (JIF) implemented by the Journal of Citation Reports (Web of Science) is the standard metrics considered in order to rank the journals of a category according their quality to the detriment of other indicators. To avoid this inconvenience, in this contribution we propose a new metric taking into account the following variables: Impact factor by JCR and by SJR, JIF 5 years, SNIP, Cite score, Total citations and Eigenfactor score. The methodology consists of summarizing all this information in two factors, using a Principal Component Analysis with VARIMAX rotation, explaining a high percentage of the total variance.

This procedure is then applied to journals included in the Odontology category where the first two factors of the rotated model explain 93.65% of the total variance.

Keywords: Factor analysis, principal component, Dentistry, Journal impact factor.

The Distribution of the Amount of Rain. A Renewall Stochastic Process Model

Javier Villarroel[1], Juan A. Vega[2]

[1]Department of Statistics and Inst. Univ. Fis. Matem., Universidad de Salamanca, Salamanca, Spain, [2]Inst. Univ. Fis. Matem., Universidad de Salamanca, Salamanca, Spain

With this work we intend to study the distribution of the amount of rain and other related statistics in the environment of Lake of Sanabria (Spain). To do this we will develop a model using a compound renewal process in which we will not assume that the initial time is a renewal time. We will deduce the resulting integral equations and solve them for various distributions of both, waiting times and jumps. We will then apply them for the study of the distribution of the amount of rain and the remaining statistics.

Keywords: Compound renewal process, non- renewal time, amount of rain, Lake of Sanabria.

A decomposable multi-type Bellman-Harris branching process with multiple supercritical types

Kaloyan Vitanov[1], M. Slavtchova-Bojkova[2]

[1]Sofia University "St. Kl. Ohridski", Bulgarian Academy of Sciences, Sofia, Bulgaria, [2]Sofia University "St. Kl. Ohridski", Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

Spread of cancer metastasis is the ultimate cause of death for most cancer patients. In order to model the complex evolution of the disease we propose a decomposable multi-type Bellman-Harris branching process. A distinctive feature of the model is that it can accommodate multiple supercritical types (of cancer cells), which in turn are used to model the

rise of new metastasis that are outside the scope of an administered treatment. As the population of a supercritical type in time approaches, theoretically, indefinite growth with a non-zero probability, it is important to study additional related characteristics. Such characteristics are the probabilities of extinction, the number of occurred supercritical mutants per type, the time until the first occurrence of a mutant starting a non-extinction process and the immediate risk for the process to escape extinction. We further discuss numerical schemes and algorithms for solving the systems of integral equations that we obtai! n.

Keywords: Mutations, Decomposable multi-type branching process, Probability of extinction, Waiting time to escape mutant, Immediate risk of escaping extinction.

Acknowledgements: The research of these authors was supported, in part, by the Bulgarian National Science Fund at the Ministry of Science and Education, Grant No. KP-06-H22/3 for theoretical development of the stochastic modelling and by the project BG05M2OP001-1.001-0004 (UNITE) funded by Operational Program Science and Education for Smart Growth co-funded by European Regional Development Fund for computational resources.

Alcohol-related mortality and years of life lost due to alcohol in Europe: an international comparison

Jana Vrabcová[1], Markéta Pechholdová[2]

[1,2]Department of Demography, University of Economics in Prague, Prague, Czech Republic

Notable differences in levels and patterns of alcohol-related harm persist within Europe. Cultural habits in alcohol consumption were identified as one of the reasons for this diversity. Traditionally, Eastern Europeans tend to drink in a more harmful way than Western Europeans. This has been reflected in higher levels of alcohol-related mortality in the East. However, times do change and the transition to free-market economies has changed the old habits, including the alcohol consumption structure and drinking attitudes. As a result, we are witnessing a unifying trend across Europe. To identify the actual size and timing of this process, we analyze long-term

cause-of-death data for a subset of European countries. We include an analysis of alcohol-specific causes of death and age patterns. Additionally, we enumerate years of life lost due to alcohol, a metric useful in the context of different mortality backgrounds. The overall aim is to assess the dynamics of alcohol-related harm, to identify the breakpoints and to detect trends using validated long-term data series from the Human Cause-of-Death database (www.causesofdeath.org).
Keywords: alcohol, mortality, years of life lost, Europe, causes of death.

Branching Random Walks and Their Applications

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

Branching random walks are crucially useful in investigations of stochastic systems with birth, death and migration of their elements. The principal attention will be paid to the properties of branching random walks on multidimensional lattices. We will be mainly interested in the problems related to the limiting behavior of branching random walks such as existence of phase transitions under change of various parameters, the properties of the limiting distribution of the particle population, existence and the shape of the propagating fronts of particles, etc. The answer to these and other questions heavily depends on numerous factors which affect the properties of a branching random walk. We describe, how the properties of a branching walk depend on the fact of non-homogeneity of the branching media, on the number and mutual disposition of the branching sources, and also on such properties of a branching walk as its symmetry and finiteness or infiniteness of the variance of jumps. We present also some results of simulation of branching random walks and discuss how they may be applied to numerical estimation of various characteristics describing the properties of the phase transitions.
The research was supported by the Russian Foundation for the Basic Research (RFBR), project No. 20-01-00487.
Keywords: Branching random walk, Multidimensional lattices, Homogeneity and non-homogeneity of environments, Limit theorems.

The Implementation of Hierarchical Classifications and Cochran's Rule in the Analysis of Social Data

Aggeliki Yfanti[1], Catherine Michalopoulou[2]

[1]Ph.D., Adjunct Lecturer, Department of Social Anthropology, Panteion University of Social and Political Sciences, Athens, Greece, [2]Professor of Statistics, Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece

In social sample survey research and the census, international organizations have developed classifications for the measurement of background variables such as the level of educational attainment (ISCED), economic activities (ISIC) and occupations (ISCO) to ensure the cross-national and overtime comparability of measurement. In this context, Eurostat developed the Nomenclature des Unités territoriales statistique (NUTS) and the European socio-economic classification (ESeC). All these classifications ─with the exception of ISCED─ are categorical schemas, defined hierarchically to allow for different levels of detail in the analyses. However, when performing bivariate analyses with these classifications, Cochran's well-known rule of thumb of expected values larger than five in order to use the chi-square test has to be taken also into account in presenting the results. In this paper, we investigate the implementation of ESeC to regions (NUTS) to demonstrate how to decide statistically on the appropriate level of classifications to be used in the analysis of social data. The analysis is based on the 2016 European Social Survey (ESS) datasets for five European countries: Austria, Belgium, France, Ireland and Italy.

Keywords: ESeC, NUTS, Cochran's rule, chi-square test, contingency table.

Dynamic Optimization with Tempered Stable Subordinators for Modeling River Hydraulics

Hidekazu Yoshioka[1], Yumi Yoshioka[2]
[1,2]Graduate School of Natural Science and Technology, Shimane University, Matsue, Japan

Lévy processes having infinite activities serve as building blocks for modern mathematical modeling and analysis of stochastic dynamical systems. In particular, tempered stable subordinators have recently been found to be right candidates for describing hydraulic processes occurring in river environment. However, their applicability and limitation have not been well-studied in environmental engineering and related research areas. In this contribution, we firstly present an exactly-solvable risk-minimizing stochastic differential game for flood management in rivers. The streamflow dynamics follow stochastic differential equations driven by a tempered stable subordinator. An entropic dynamic risk measure is employed to evaluate a flood risk under model uncertainty. The problem is solved via a Hamilton–Jacobi–Bellman–Isaacs equation. We explicitly derive an optimal flood mitigation policy of a hydraulic structure along with the worst-case probability measure of river flows. A related backward stochastic differential equation for water quality dynamics is also briefly discussed with demonstrative numerical computation using a data collected at a river.

Keywords: Tempered stable subordinators, Backward stochastic differential equations, Flow discharge, Water quality dynamics, Dynamic risk measures.

Investigating the current state of population health and mortality in Niger

Ibrahim Sidi Zakari
Dpt of Mathematics and Computer Science, Abdou Moumouni University, Niamey, République du Niger

This paper aims at investigating data and databases related to Nigerien population health state and mortality. Fact that will contribute to identify potential data gaps in the context of the United Nations Sustaible Development Goals (agenda 2030) as well as new challenges and opporunities in terms of definition, modeling and estimates.

Spreading Disease modeling using Markov Random Fields

Stelios Zimeras
University of the Aegean, Department of Statistics and Actuarial - Financial Mathematics, Karlovassi, Samos, Greece

In recent years stochastic models and statistical methods have been successfully applied in image analysis. Of particular interest are Bayesian methods based on local characteristics. The use of Bayesian methods is an approach, which seeks to provide a unified framework in modeling within many different image processes. Models based on Markov random fields are widely used to model spatial processes. Key components of any statistical analysis using such models are the choice of an appropriate model as the prior distribution and the estimation of prior model parameters. Models for spreading diseases are given based on whether or not the disease succeeds or fails to appear in the region. For this reason, a logistic model could be proposed, with white patches representing success and black representing failure. For that reason, a logistic / binomial model could be proposed with coloring white in success and black in the failure. Due to the 2D definition of the region and the conditionality of the pixel values, these particular models are defined as auto-models. In practice estimates using Bayesian methods cannot be computed analytically. For this reason Monte Carlo algorithms can be

used to generate samples from the posterior distribution and parameter estimates calculated from this sample. The fundamental idea is to use an algorithm, which generates a discrete time Markov chain converging to the desired distribution. The most commonly algorithms include the Gibbs sampler and the Metropolis-Hasting algorithm. In this work, the spatial pattern models for spreading diseases have been analyzed considering Markov random fields auto-models. The Gibbs sampler would be used to simulate example images for various parameter combinations

Keywords: Spreading Disease, Markov random fields, Bayesian analysis, estimation techniques, Auto-logistic, Auto-binomial models.

Convergence of Compound Renewal Processes in a Heavy-tailed Set-up

Nadiia Zinchenko

Department of Information Technology and Data Analysis, Nizhyn State Mukola Gogol University, Nizhyn, Ukraine

In this work we present a number of strong and weak limit theorems for the compound renewal process D(t) = S(N(t)), where N(t) is a renewal counting process and S(·) is an usual partial sum process, under various assumptions on renewal process, moment and dependent conditions on random summands $\{x_i, i \geq 1\}$. A lot of attention is paid to the case of heavy-tailed summands, particularly, when $\{x_i, i \geq 1\}$ are attracted to α -stable law. The rate of growth of the process D(t) and its increments D(t+a(t)) - D(t), when a(t) grows as $t \to \infty$, is studied in details. Various applications in risk theory, financial analysis and queuing theory are discussed with emphasis on actuarial applications (asymptotic behavior of risk processes with large claims).

Keywords: Renewal Process, Compound Renewal Process, Limit Theorems, Rate of growth, Risk Theory, Risk Process.