StatMod2020

Statistical Modeling with Applications

November 6-7, 2020

"Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania

---

Organizers:

"Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania

and

Laboratory of Mathematics Raphael Salem, University of Rouen-Normandy, France

---

ABSTRACTS

Alexandru AGAPIE (Bucharest Univ. of Economic Studies, Romania)
Evolutionary algorithms on continuous space. Uniform mutation inside the sphere
joint work with Ovidiu SOLOMON (Bucharest Univ. of Economic Studies & Institute of Solid Mechanics, Romania)

Abstract: We tackle the problem of analytically describing the local probability transition kernel of a continuous Evolutionary Algorithm. Instead of the usual normal mutation, we propose the uniform distribution inside the sphere. This type of distribution has barely been addressed in previous studies. The study is an in depth analysis of the subject, and delivers remarkable results, concerning both the analytic tractability, and the comparison against the algorithm with normal mutation.

Bogdan ALEXE ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics & Univ. of Bucharest, Romania)
Does the lack of supporters affect the home team advantage in football matches? The COVID-19 pandemic case.

joint work with Denis ENACHESCU ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)

Abstract: The home team advantage plays an important role in the performance of a football team. Statistically, teams playing at home score more goals and concede less, consequently, earning more points and performing better than teams playing away. One of the many factors influencing the home team advantage is the support of the home crowd. The COVID-19 pandemic has changed the conditions in which football matches are played with supporters being partially or totally banned in attending football matches. In this paper we try to give a statistical answer if the lack of supporters affects the home team advantage in football matches. We based our studies on data containing football matches from the most known European leagues (England, Spain, Germany, Italy) played before and after the COVID-19 pandemic emerged.

Chafiàa AYHAR (LSMSA, Univ. Dr. Moulay Tahar of Saïda, Algeria)
Reliability of semi-Markov systems
joint work with Fatiha MOKHTARI (LSMSA, Univ. Dr. Moulay Tahar of Saïda, Algeria), Saâdia RAHMANI (LSMSA, Univ. Dr. Moulay Tahar of Saïda, Algeria), Vlad Stefan BARBU (LMRS, Univ. of Rouen-Normandy, France)

Abstract: We consider a semi-Markov process, with a finite state space. We review the empirical and nonparametric estimation of reliability, availability and failure rate. We also give asymptotic properties of the above estimators, as the uniform strong consistency and normality. This paper is an overview of some result of Ouhbi and Limnios (1999, 2003).

Luiza BĂDIN (Bucharest Univ. of Economic Studies & "Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
An integrative framework for evaluating scientific research
joint work with Silvia DEDU, Anca-Teodora ȘERBAN OPRESCU, Florentin ȘERBAN (Bucharest Univ. of Economic Studies, Romania)

Abstract: Research is a very particular and unconventional type of activity for which a hidden incentive is the academic freedom. Although many voices claim that the concept of efficiency has nothing to do with research per se, recent results on performance evaluation in research and development show that advanced econometrics and operational research methodologies are extremely useful and relevant in analyzing the efficiency and productivity of different sectors and activities, including scientific research. The aim of the paper is to analyze and compare from an interdisciplinary perspective some of the latest approaches, methodologies and tools for evaluating the

efficiency of scientific research. The paper contributes to the theoretical development and practical implementation of the most recent techniques of efficiency measurement for performance assessment and ranking of scientific research units.

Ionuț BEBU (George Washington Univ., USA)
Statistical considerations in studies with composite outcomes

Abstract: Many clinical studies (e.g., cardiovascular outcome trials) investigate the effect of an intervention on multiple event-time outcomes. The most common method of analysis is a so-called "composite" analysis of a composite outcome defined as the time to the first component event. Other approaches have been proposed, including the win ratio (or win difference) for ordered outcomes and the application of the Wei-Lachin test. We provide a brief overview of these methods, along with an extension to a severity-weighted win-ratio approach. Data from the Prevention of Events with Angiotensin Converting Enzyme Inhibition (PEACE) clinical trial is used as illustration.

Bogdan BIOLAN (Univ. of Bucharest, Romania)
About iterative approaches for sparse direction-of-arrival estimation for co-prime arrays with off-grid targets

Abstract: This paper is about the problem of direction of arrival (DOA) estimation by the link with the sparsity enforced recovery technique for co-prime arrays, which is useful to increase the degrees of freedom. In order to apply the sparsity based technique, the discretization of the potential DOA range is required. Therefore, every target must fall on a predefined grid. We introduce a convex function majorizing a constructed objective function, and an iterative approach is developed to amend gradually the offset vector to get the final DOA estimation. Numerical simulations are done in order to verify the accuracy and effectiveness of the proposed methods in terms of potential detection, resolution ability and root mean squared estimation error, as we compare with other state-of-the-art methods.

Maroua BOUBRED (LSMSA, Univ. Dr. Moulay Tahar of Saïda, Algeria)
Semi-Markov model applied to Algeria Coronavirus data
joint work with Fatiha MOKHTARI (LSMSA, Univ. Dr. Moulay Tahar of Saïda, Algeria)

Abstract: In this work, we explained the continuous-time semi-Markov model with a discrete set of states. We defined empirical estimators of important quantities such as semi-Markov kernel, sojourn time distributions, transition probabilities, and hazard rate function. We gave results about their asymptotic properties.

The present work aims at the introduction of the continuous-time semi-Markov model as a candidate model for the description of Algeria Coronavirus data.

The process of Algeria Coronavirus data was represented with the parametric and nonparametric methods. Semi-Markov package in R Language was used for the implementation of the parametric-method however, for the nonparametric one, we had developed our functions.

Badreddine BOUMARAF (Univ. of Souk-Ahras, Algeria & Univ. of Badji Mokhtar of Annaba, Algeria & LMRS, Univ. of Rouen-Normandy, France)
Estimation of Beta-Pareto distribution based on several optimization methods
joint work with Nacira SEDDIK-AMEUR (Univ. of Badji Mokhtar of Annaba, Algeria), Vlad Stefan BARBU (LMRS, Univ. of Rouen-Normandy, France)

Abstract: This paper is concerned with the maximum likelihood estimators of the Beta-Pareto distribution. Since these estimators cannot be obtained explicitly, we use nonlinear optimization methods that numerically provide these estimators. The methods we investigate are the method of Newton-Raphson, the gradient method and the conjugate gradient method. In order to compare between several concurrent models, namely generalized Beta-Pareto, Beta, Pareto, Gamma and Beta-Pareto, model criteria selection are used. We consider completely observed data, as well as right censored. This presentation is based on Boumaraf, Seddik-Ameur, Barbu (2020).

Luigi-Ionuț CATANĂ (Univ. of Bucharest, Romania)
Stochastic order of a multivariate uniform distributions family
joint work with Anişoara RĂDUCAN ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics)

Abstract: We give sufficient or necessary conditions of stochastic order for a multivariate uniform distributions family.

Simona COJOCEA (Univ. of Bucharest, Romania)
On the variable selection property of the Lasso method when p >> n

Abstract: Regularization methods are commonly used when we need to introduce a set of constraints in the data, when fitting a model. The tuning parameter is allowing us to have some control over the estimated coefficients, which we generally want to shrink. While Ridge regression is very successful at shrinking the coefficients but keeping their absolute value strictly positive, the Lasso manages to shrink down to 0 some of the coefficients, as the value of the tuning parameter increases. This way, the Lasso is performing variable selection, which is particularly useful in a high-dimensional setup.

In this paper we discuss the variable selection property of the Lasso when p>>n, we assess its performance, and we try to derive some conclusions on the possible interpretation of the model we obtain, having as a starting point a case study from genetics.

Doru CONSTANTIN (Univ. of Pitești, Romania)
Applications of extended cumulative entropies in image processing
joint work with Costel BĂLCĂU (Univ. of Pitești, Romania), Vasile PREDA ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics & "Costin C. Kiritescu" National Institute of Economic Research & Univ. of Bucharest, Romania)

Abstract: Several extended cumulative entropies have been proposed in recent years. We present some applications of these entropies in image processing fields.

Guglielmo D'AMICO (Univ. "G. d'Annunzio" of Chieti-Pescara, Italy)
On the computation of some interval reliability indicators for semi-Markov systems
joint work with Raimondo MANCA (Univ. of Rome, "La Sapienza", Italy), Filippo PETRONI (Marche Polytechnic Univ., Italy), Selvamuthu DHARMARAJA (Indian Institute of Technology Delhi, India)

Abstract: In this paper, we compute general interval indicators of availability and reliability for time inhomogeneous semi-Markov chains. First, we consider duration dependent extensions of the Interval Reliability and then, we determine an explicit formula for the availability with a given window and containing a given point. The window availability can be computed through an explicit formula involving duration-dependent transition probabilities and the interval reliability function. Both interval reliability and availability functions are evaluated considering the local behavior of the system through the recurrence time processes. The results are illustrated through a numerical example. They show that the considered indicators can describe the duration effects as well as the age of the multi-state system and could be effectively applied to real-life problems.

Bice DI BASILIO (Univ. "G. d'Annunzio" of Chieti-Pescara, Italy)
A semi-Markovian approach to drawdown based measures
joint work with Guglielmo D'AMICO (Univ. "G. d'Annunzio" of Chieti-Pescara, Italy), Filippo PETRONI (Marche Polytechnic Univ., Italy)

Abstract: Risk is an essential concept that must be known and understood by all savers, in order to make informed decisions. It is an inherent feature of financial assets that

expresses uncertainty about the future value of a security and it should always be kept in mind while operating in the stock markets. Typically, it manifests in continuous movements in prices, upward or downward, able to generate crashes of different intensity, affecting stakeholder' earnings. As a consequence, the main investor's goal is to quantify risk by means of risk measures, avoiding suffering heavy portfolio losses. Over the years a vast and varied literature on this topic has developed. In this paper we focus on risk indicators based on drawdown and closely related to market collapses: the drawdown of fixed level, the time to crash, the speed of crash, the recovery time and the speed of recovery. In detail the drawdown of fixed level, the time to crash and the recovery time scan the time in which the drawdown of an asset attains a selected K-level for the first time, the time taken to have the first K-variation and the time necessary to have the first K'-descendent after having the first K-ascent, respectively. They provide extremely valuable information in portfolio management, especially in selecting and controlling a group of investments. In order to analyze these measures, we use intra-day prices of Fiat stock, quoted on the Italian stock exchange, and we apply to it the weighted-indexed semi-Markov chain (WISMC) model. It is an enhancement of the general SMC model that allows to repeat the long dependence structure in the squared returns by adding an index process. To test the validity of our model we realize comparisons with the GARCH and EGARCH models, widely used in finance. It should also be emphasized that our application is performed considering the censored data in the estimation procedures. Globally, the WISMC model performs better than the selected GARCH and EGARCH models, for the time to crash. As for the recovery time, our model is more efficient than the others only if the percentage of censored units is less than 30%.

Mihăiță DRĂGAN (Univ. of Bucharest, Romania)
Neural networks classified using Tsallis entropy

Abstract: Since the beginning of the new millennium, increased attention has been paid to entropy-based assessment criteria in adaptive systems. Thus, for example, for the classification of complex networks (eg. neural networks), several principles have been proposed based on maximizing or minimizing entropic cost functions, starting from applying Shannon's entropy. Unlike the conventional graph theory, complex networks contain an enormous number of nodes (up to a few million) and a complex topology with irregular interconnection.

Therefore, Shannon entropy used in standard decision trees, from top to bottom, does not guarantee the best generalization. Separate criteria based on generalized entropy offer a different compromise between node purity and general information gain. Under these conditions, a comparative study based on Tsallis entropy was made from the

results obtained by applying Shannon entropy, so this approach can be used in any decision-making as well as in the information selection algorithm.

Denis ENĂCHESCU ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
Monte Carlo methods for solving parabolic partial differential equations

Abstract: We formulate the solution of the parabolic problem as a limit of a Neumann series and built a Markov chain to estimate the solution of the PDE for a given point.

We investigate the approximation error and the convergence and study the efficiency of the method in terms of the selection volume necessary to obtain a given error and the average number of operations.

Finally, we present some numerical results comparatively with other methods.

Thomas GKELSINIS (LMRS, Univ. of Rouen-Normandy, France)
Theoretical aspects and practical implications on measures of weighted information
joint work with Vlad Stefan BARBU (LMRS, Univ. of Rouen-Normandy, France), Alex KARAGRIGORIOU (Univ. of the Aegean, Greece)

Abstract: Measures of weighted information are obtained through classical measures of information by taking into account the importance of specific qualitative characteristics of each event. These measures are classified into two main categories, the entropic and the divergence measures. In risk analysis as well as in theory of rare events and in reliability theory it is common to take under consideration for statistical purposes, the existence of fat (or light) tails in the distribution under investigation (most commonly the left tail). Such purposes include among others the inference based exclusively either on the tail part of the distribution or on the main part of the distribution by trimming extreme values. Motivated by these needs we present in this work, measures of directed information. The proposed measures quantify the underlying information with emphasis on specific parts (or events) of their probability distribution, without losing the information of the less significant parts.

Steven GOLOVKINE (ENSAI Rennes & CREST, France)
Multivariate functional data clustering using unsupervised binary tree
joint work with Nicolas KLUTCHNIKOFF (IRMAR, Univ. Rennes 2, France) and Valentin PATILEA (ENSAI Rennes & CREST, France)

Abstract: With the recent development of sensing devices, more and more data are recorded in both dimensions of time and space. These measures lead to large amounts of

data that are often referred as multivariate functional data. This work proposes a simple clustering procedure for such multivariate functional data. Considering a multivariate functional principal components analysis as a dimension reduction vehicle, a binary tree is grown using a parametric mixture model defined on the projection of the trajectories onto the principal components. The mixture model is fitted by an EM algorithm. Then, a joining step is introduced to eventually merge the similar nodes of the tree that do not share a direct ascendant. A detailed description of the algorithm is provided, along with an extensive numerical analysis on both simulated and real datasets.

Rokas GYLYS (Vilnius Univ., Lithuania)
Modelling uncertainty inhuman mortality projections using state-space Lee-Carter

Abstract:

Lamia HAMMADI (LabSIPE, ENSA of El Jadida, Morocco)
Hidden Markov models for risk time series in customs supply chain
joint work with Vlad Stefan BARBU (LMRS, Univ. of Rouen–Normandy, France), Eduardo SOUZA DE CURSI (LMN, INSA of Rouen, France

Abstract: In this paper, Hidden Markov Models (HMMs) are proposed to describe risk time series in customs supply chain context. We start by introducing the properties of the Markov processes and providing a rigorous presentation of the main problems involved with Hidden Markov Models. Then we discus an extension where the hidden Markov chain is non-stationary to deal with seasonal fluctuations of risks in customs context. The different models are fitted using the EM-algorithm to time series of quantities seized of the illicit traffic on two sites. It is shown that the fitted models can be easily interpreted and provide a good description of important properties of the data such as the second-order structure or estimate the total illicit traffic per site.

Emmanouil-Nektarios KALLIGERIS (Univ. of the Aegean, Greece)
Modeling time-series incidence data via a multiple changepoint approach
joint work with Alex KARAGRIGORIOU (Univ. of the Aegean, Greece), Christina PARPOULA (Panteion Univ. of Social and Political Sciences, Greece)

Abstract: Modeling of incidence data, has been in the center of interest in various scientific fields such as epidemiology, medicine, meteorology, etc. Several attempts have been made over the years in order to identify mainly the typical periods of such data, and there exists a variety of techniques to achieve so. The goal of this work is the modeling of both typical and non-typical periods that occur in time-series incidence data.

The identification of non-typical periods is achieved by using changepoint detection analysis. In addition, model selection techniques are implemented in order to identify the optimal Periodic-type Auto-Regressive Moving Average model with covariates that best describes the pattern of the time-series. Finally, an advanced algorithm was developed in order to improve the estimation accuracy of the selected model and tested in both real and simulated datasets, which were based on a real dataset concerning Influenza-like illness rate for Greece for the period 2014-2016, with satisfactory results.

Maria LESHCHINSKAIA (Tomsk State Univ., Russia)
Efficient estimation of non-parametric signals observed with the Levy noises of small intensity
joint work with Evgenii PCHELINTEV (Tomsk State Univ., Russia), Serguei PERGAMENCHTCHIKOV (LMRS, Univ. of Rouen–Normandy, France)

Abstract: We develop efficient non parametric estimation theory for continuous time regression models with non-Gaussian Lévy noises in the case when unknown functions belong to the Sobolev ellipse. On the basis of the approach proposed by Pinsker in 1981 we provide sharp lower bounds for the asymptotic mean square accuracy. It turns out that the main result obtained by Pinsker for the "signal plus white noise" model is not correct without additional conditions for the ellipse. In this paper we find constructive sufficient conditions for the ellipse under which we construct efficient estimators. In particular, we show that the obtained conditions hold for the coefficient of an exponential form. It should be emphasized that in this talk for the first time we calculate the sharp lower bound in explicit form for exponential coefficients. Finally, we apply this result to the estimation problem for the number of signals in multi-pass connection channels and, as consequence, we obtain a natural for this case almost parametric convergence rate which considerably improves the rate with respect to the usually used ellipse power coefficients.

Oana LUPAŞCU STAMATE ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
Numerical approach for stochastic differential equations of fragmentation; application to avalanches
joint work with Lucian BEZNEA (Simion Stoilow Institute of Mathematics of the Romanian Academy & Univ. of Bucharest & Centre Francophone en Mathématiques de Bucharest, Romania), Madalina DEACONU (IECL, Univ. of Lorraine, CNRS & Inria, France)

Abstract: We consider a stochastic fragmentation process of an infinite particle

system. We introduce a stochastic differential equation of fragmentation with either a continuous or a discontinuous fragmentation kernel and we give several examples. As an application, we consider a stochastic model for the fragmentation phase of an avalanche, involving a physical fractal property. Finally, we complete the theoretical approach with numerical simulations.

Bojana MILOŠEVIĆ (Univ. of Belgrade, Serbia)

On characterization based goodness-of-fit tests: a review of recent results

Abstract: Recently, characterization based approach for the construction of goodness-of-fit tests has become rather popular. Such tests are attractive because they reflect some intrinsic properties of probability distributions connected with the given characterization, and therefore can be more efficient or more robust than others. The growth of the number of characterization theorems especially contributed to the development of this direction.

In this talk, we present several types of characterization theorems as well as several approaches for the construction of goodness-of-fit tests. General approaches will be accompanied with particular examples of recently proposed tests for Exponential, Pareto, Uniform and Lévy distribution. The presentation will be based on the references given below, as well as some working results. In addition, some promising directions for future research will be given.

References

[1] M. Cuparíc, B. Miloševíc, Y. Y. Nikitin, and M. Obradovíc. Some consistent exponentiality tests based on Puri-Rubin and Desu characterizations. *Applications of Mathematics*, 65(3):245–255, 2020.

[2] M. Cuparíc, B. Miloševíc, and M. Obradovíc. New consistent exponentiality tests based on V-empirical Laplace transforms with comparison of efficiencies. *arXiv preprint arXiv:1904.00840*, 2019.

[3] M. Cuparíc, B. Miloševíc, and M. Obradovíc. New L^2-type exponentiality tests. *SORT*, 43(1):25–50, 2019.

[4] M. Jiḿenez-Gamero, B. Miloševíc, and M. Obradovíc. Exponentiality tests based on Basu characterization. *Statistics*, pages 714–736, 2020.

[5] B. Miloševic. Some recent characterization based goodness of fit tests. In *the Proceedings of the 20th European Young Statisticians Meeting*, pages 67–73, 2017.

[6] B. Miloševíc. Asymptotic efficiency of goodness-of-fit tests based on Too-Lin charac-

terization. *Communications in Statistics-Simulation and Computation*, 49(8):2082– 2101, 2020.

[7] B. Miloševi´c and M. Obradovi´c. Two-dimensional Kolmogorov-type goodness-of-fit tests based on characterisations and their asymptotic efficiencies. *Journal of Nonparametric Statistics*, 28(2):413–427, 2016.

Andreas MAKRIDES (Univ. of Uclan, Cyprus & Univ. of the Aegean, Greece)
Statistical inference in models with parameter dependence
joint work with Vlad Stefan BARBU (LMRS, Univ. of Rouen-Normandy, France),
Alex KARAGRIGORIOU (Univ. of the Aegean, Greece)
Abstract: In this work we are interested in a general class of distributions for independent not necessarily identically distributed (inid) random variables, closed under extrema, that includes a number of discrete and continuous distributions like the Geometric, Exponential, Weibull or Pareto. The scale parameter involved in this class of distributions is assumed to be time varying with several possible modeling options proposed. Such a modelling setting is of particular interest in reliability and survival analysis for describing the time to event or failure. The maximum likelihood estimation of the parameters is addressed, and the asymptotic properties of the estimators are discussed. We provide real and simulated examples and we explore the accuracy of the estimating procedure as well as the performance of classical model selection criteria in choosing the correct model among a number of competing models for the time-varying parameters of interest. This presentation is based on Barbu, Karagrigoriou, Makrides (2020).

Christina PARPOULA (Panteion Univ. of Social and Political Sciences, Greece)
A distribution-free multiple change-point model for Phase I analysis of individual observations
joint work with Alex KARAGRIGORIOU (Univ. of the Aegean, Greece)

Abstract: A distribution-free methodology for Phase I statistical process control (SPC) of individual observations is developed. One of the most common challenges in nonmanufacturing control chart applications is that the parametric form of the process distribution is unavailable. In the literature, most of the existing non-parametric control charts are designed for Phase II monitoring, and those designed for Phase I analysis can be applied with subgrouped observations [2]. In this work, we develop an alternative framework for SPC of individual observations, designed for Phase I analysis, by adopting a multiple change-point model and a permutation approach to hypothesis testing [1]. Under this framework, a distribution-free control chart which works as a mean shift detector is constructed. The in-control and out-of-control performances of the

developed control scheme are thoroughly investigated, for normal and nonnormal observations, through an extensive Monte Carlo simulation study, comparing the developed chart with some representative parametric and nonparametric counterparts existing in the literature. The performance of the proposed chart, in terms of the signal probability, compares favorably with recently developed charts (based on generalized likelihood ratio with permutation based control limits) for individual observations [2]. A real-life example is given to illustrate the design and implementation of the proposed scheme in practice, and to unfold its capabilities. Some empirical guidelines are provided for practitioners to choose the most appropriate control chart under different distributions and shift patterns. Concluding remarks and suggestions for future research are given.

References

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

[2] Parpoula C, Karagrigoriou A. On change-point analysis-based distribution-free control charts with Phase I applications. In Koutras M, Triantafyllou I, eds. Distribution-free Methods for Statistical Process Monitoring and Control. Springer International Publishing; 2020; 157-182.

Evgenii PCHELINTEV (Tomsk State Univ., Russia)

Improved estimation methods for the conditional Gaussian Ornstein–Uhlenbeck non parametric regression models based on discrete data

joint work with Serguei PERGAMENCHTCHIKOV (LMRS, Univ. of Rouen–Normandy, France), Maria LESHCHINSKAIA (Tomsk State Univ., Russia)

Abstract: We study a nonparametric regression model observed in the discrete time moments with noises defined through Ornstein–Uhlenbeck processes with jumps. For this model improved (shrinkage) estimation methods are developed and the non-asymptotic comparison between shrinkage and least squares estimates are studied. The improvement effect for the shrinkage estimates showing the significant advantage with respect to the parametric models is established. It turns out that obtained improvement effect holds true uniformly over observation frequency. Then, through these shrinkage estimators an adaptive model selection procedure is constructed and non-asymptotic sharp oracle inequalities for this procedure are obtained. Constructive sufficient conditions for the observation frequency providing the robust efficiency property in adaptive setting are found. Monte-Carlo simulations for the numeric confirmation of the obtained theoretical results are given

Serguei PERGAMENCHTCHIKOV (LMRS, Univ. of Rouen–Normandy, France)
Adaptive efficient robust sequential estimation for high dimensional autoregressive models
joint work with Ouerdia ARKOUN (Sup'Biotech & LMRS, Univ. of Rouen–Normandy, France), Jean-Yves BRUA (LMRS, Univ. of Rouen–Normandy, France)

Abstract: In this paper we study high dimension models based on dependent observations defined through non parametric autoregressive processes. For such models we study the efficient robust sequential estimation problem in adaptive settings, i.e. in the case when the nonparametric regularity is unknown. To this end we use the model selection procedures proposed by Arkoun, Brua and Pergamenchtchikov (2019). First, through the Van Trees inequality, we obtain the sharp lower bound for robust risks in explicit form, i.e. the famous Pinsker's constant. Then, through sharp non asymptotic oracle inequalities for robust risks, we show that the upper bound for the robust risk of the proposed model selection sequential procedure coincides with the obtained Pinsker constant, i.e. this means that this procedure is efficient in the minimax sense.

Filippo PETRONI (Marche Polytechnic Univ., Italy)
Actuarial methods for hedging wind power variability
joint work with Guglielmo D'AMICO (Univ. "G. d'Annunzio" of Chieti-Pescara, Italy), Fulvio GISMONDI ("Guglielmo Marconi" Univ., Italy)

Abstract: The share of production due to wind is continuously increasing in time although there are still relevant problems that affect this industry. The most important limitation for a further development of the wind energy industry concerns the variability of the wind speed phenomenon. The problem of the wind speed volatility has been approached mainly by energy storage systems; that is, by storing a surplus of energy to be used for compensating an eventual future deficit of production. More recently an insurance contract between the wind energy producer and a dispatchable energy producer has been proposed as a mean to manage the uncertainty of the wind speed. In this paper we extend previous results involving the use of insurance contracts by considering the dependence existing between electricity prices and wind energy production. The dependence structure is modelled using an appropriate copula function and we show the impact of this dependence on the fair premium that the wind power supplier has to pay in order to hedge the risk of inadequate output of electricity at any time. Recursive type equations are obtained for the prospective mathematical reserves of the insurance contract and for their higher order moments. The model and the validity of the results are illustrated through a numerical example.

Eugen PIRCALABELU (UC Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences, Belgium)
Community detection on probabilistic graphical models with group-based penalties
joint work with Gerda CLAESKENS (KU Leuven, ORSTAT & Leuven Statistics Research Center, Belgium)

Abstract: A new strategy of probabilistic graphical modeling is developed that draws parallels from social network analysis. Probabilistic graphical modeling summarizes the information coming from multivariate data in a graphical format where nodes, corresponding to random variables, are linked by edges that indicate dependence relations between the nodes. The purpose is to estimate the structure of the graph (which nodes connect to which other nodes) when data at the nodes are available. On the opposite side of the spectrum, social network analysis considers the graph as the observed data. Given thus the graph where connections between nodes are observed rather than estimated, social network analysis estimates models that represent well an underlying mechanism which has generated the observed graph.

We propose a new method that exploits the strong points of each framework as it estimates jointly an undirected graph and communities of homogenous nodes, such that the structure of the communities is taken into account when estimating the graph and conversely, the structure of the graph is accounted for when estimating homogeneous communities of nodes. The procedure uses a joint group graphical lasso approach with community detection-based grouping, such that some groups of edges co-occur in the estimated graph. The grouping structure is unknown and is estimated based on community detection algorithms.

Theoretical derivations regarding graph convergence and sparsistency, as well as accuracy of community recovery are included, while the method's empirical performance is illustrated in an fMRI context, as well as with simulated examples.

Cristian PREDA (Univ. of Lille, France & "Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
One dimensional discrete scan statistics for dependent models and some related problems
joint work with Alexandru AMĂRIOAREI (Univ. of Bucharest, Romania)

Abstract: The one dimensional discrete scan statistic is considered over sequences of

random variables generated by block factor dependence models. Viewed as a maximum of an 1-dependent stationary sequence, the scan statistics distribution is approximated with accuracy and sharp bounds are provided. The longest increasing run statistics is related to the scan statistics and its distribution is studied. The moving average process is a particular case of block factor and the distribution of the associated scan statistics is approximated. Numerical results are presented.

Marius RĂDULESCU ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
Single period portfolio selection models with transaction costs and initial holdings
joint work with Constanța Zoie RĂDULESCU (National Institute for Research and Development in Informatics, Romania)

Abstract: The classical mean-variance model in the portfolio theory does not include transaction costs and initial holdings. In this paper we extend Markowitz's portfolio selection model to include transaction costs in the presence of initial holdings for the investor. Our approach is new. Our aim is to obtain an optimal portfolio which has a minimum risk or a maximum return. The optimal solution may require both buying and selling a particular asset, which is clearly not an advisable practical strategy. In order to have a good strategy it is necessary to include in the portfolio selection models complementarity constraints. These constraints do not allow that the same asset be bought and sold at the same time. That is why our portfolio selection models include complementarity constraints. This type of constraints increases the difficulty of the problems, which now enter in the category of combinatorial optimization problems. The set of feasible solutions for the problems from the above mentioned class is the union of a set of convex sets but it is no longer convex. We study several approaches for finding solutions of portfolio selection models with complementarity constraints. Several numerical results are discussed.

Elena-Grațiela ROBE-VOINEA (Naval Academy "Mircea cel Bătrân", Constanța, Romania)
Solving actuarial problems with FastCalc V.0.1.

Abstract: The actuarial problem that is discussed in this paper is represented by the calculation of the distribution of the aggregate claims value of an insurance portfolio by certain methods such as: Panjer Method, Fast Fourier Transform method and Simulation using FastCalc V.0.1., a software program developed entirely in the Matlab software environment.

Mathilde SAUTREUIL (MICS, CentraleSupélec, Univ. Paris-Saclay, France)
Study of neural networks to predict the survival in oncology
joint work with Sarah LEMLER, Paul-Henry COURNÈDE (MICS, CentraleSupélec, Univ. Paris-Saclay, France)

Abstract: In this talk, we are interested in studying the potential of neural networks to predict survival in oncology. In clinical study in oncology, the number of variables to characterize patients can be very large, it can include clinical, genomic, radiology image data, while the number of patients in cohorts remains relatively small. The classical methods used until now, like the Cox model [2], do not work anymore in high-dimension, i.e. when the number of covariates is larger than the size of samples. Recently, the neural networks made a come-back in some research domains thanks to the increase of computing power. The survival analysis is one of these domains. Some neural networks have already been developed, but few of them were studied in high-dimension. Two strategies have been used to develop models based on neural networks in survival analysis. The first one is based on the Cox model and the second one is based on a discrete-time model. This second approach was less studied than the one based on the Cox model. We focus on neural networks based on a discrete-time model that we have adapted to the high-dimensional setting. In this talk, we present a comparison study to observe the impact of different models in survival analysis in the context of high-dimension. We compare a neural network based on the Cox model called Cox-nnet [1] with those based on a discrete-time model adapted to the high-dimension that we call NNsurv, NNsurv deep and NNsurvK. The Lasso procedure [4] using the Cox partial log-likelihood [3] is used as benchmark. We created a simulation plan to make this comparison more relevant. The data are simulated from different survival models (Cox, AFT and AH) to have data of different complexity levels (checking the proportional risks assumptions, survival curves crossing) with varying sample size and number of covariates. We also study the effect of censorship and sparsity. We have considered the Concordance index and the Integrated Brier Score to compare the performances of the different procedures. Finally, we use two classical real datasets for the comparison. We conclude that in the most situations, the best is the one based on the Cox framework in which the neural network is used to handle nonlinear effects as well as interactions (Cox-nnet). However, the model in which the neural networks directly predict the discrete risks, with several hidden layers (NNsurv deep), proves superior in the most complex situations, notably with non-proportional risks and crossing survival curves.

References

[1] T. Ching, X. Zhu, and L. X. Garmire. Cox-nnet: An artificial neural network method for

prognosis prediction of high-throughput omics data. PLOS Computational Biology, 14(4):e1006076, Apr. 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006076. URL https://dx.plos.org/10.1371/journal.pcbi.1006076.

[2] D. R. Cox. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological), 34(2):187{220, 1972. ISSN 0035-9246. URL https://www.jstor.org/stable/2985181.

[3] D. R. Cox. Partial likelihood. Biometrika, 62(2):269-276, Aug. 1975. ISSN 0006-3444. doi: 10.1093/biomet/62.2.269. URL https://academic.oup.com/biomet/article/62/2/269/337051.

[4] R. Tibshirani. The Lasso Method for Variable Selection in the Cox Model. Statistics in Medicine, 16(4):385-395, 1997. ISSN 1097-0258. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3.

Sorina-Cezarina SFETCU (Univ. of Bucharest, Romania)
Ordering properties concerning Tsallis quantile entropy

Abstract: We give some order results for Tsallis quantile entropy and study the closure, the reversed closure properties of them. As applications of a main theorem we investigate the preservation of this order in some stochastic models.

References

[1] V. Kumar, V. Rekha, A quantile approach of Tsallis entropy for order statistics, Physica A: Statistical Mechanics and its Applications, 503 (2018), 916-928.

[2] A.K. Nanda, P.G. Sankaran, S.M. Sunoj, Rényi's residual entropy: A quantile approach, Statist. Probab. Lett. 85 (2014), 114-121.

[3] V. Preda, C. Bălcău, Entropy optimization with applications, Editura Academiei Romane, Bucuresti, 2011.

[4] S.M. Sunoj, P.G. Sankaran, Quantile based entropy function, Statist. Probab. Lett. 82 (2012), 1049-1053.

[5] L. Yan, D.-t. Kang, Some new results on the Rényi quantile entropy ordering, Statistical Methodology 33 (2016), 55-70.

Muhammad SHERAZ (Institute of Business Administration Karachi, Pakistan)
Extensive and non-extensive entropy measures with financial applications
joint work with Vasile PREDA ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics & "Costin C. Kiritescu" National Institute of Economic Research & Univ. of Bucharest, Romania), Silvia DEDU (Bucharest Univ. of

Economic Studies, Romania)

Abstract: In this paper, we use extensive and non-extensive entropy measures to obtain risk-neutral densities (RNDs) and volatility computation of financial returns. For this purpose, we discuss the well-known Black-Scholes framework with examples of RNDs and volatility computations. On the other hand, we use the framework of entropy pricing theory to propose various RNDs based on the concepts of expected-utility and weighted-entropy. Finally, we present computations of European Call, Put and Greeks.

Yousri SLAOUI (LMA, Univ. of Poitiers-Futuroscope, France)
Functional data prediction and classification

Abstract: Within the framework of big-data, we are very often led to treat a voluminous set of data.

First, we consider stochastic algorithms to build recursive estimators. The major interest of these recursive approaches is that they allow a quick update of the data.

We then focus on the problem of recursive estimation of a regression function in the case of functional data, we present some results concerning the asymptotic behavior of the proposed non-parametric estimator, we then proposed a data-driven bandwidth selection procedures of the smoothing parameter and we compare the proposed method with existing methods using simulated data and then real data.

Moreover, we address the problem of the supervised classification of curves, we underline the gain of the use of recursive approaches using data simulated and then real data.

Finally, we consider the problem of the unsupervised classification using an application example from the field of Psychology more precisely in electroencephalography (EEG) which underlines the practical interest of the method.

Aida TOMA (Bucharest Univ. of Economic Studies & "Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
Model selection criteria based on pseudodistances
joint work with Alex KARAGRIGORIOU (Univ. of the Aegean, Greece), Paschalini TRENTOU (Univ. of the Aegean, Greece)

Abstract: 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. The case of the linear regression models is studied and a specific pseudodistance based criterion is proposed. Examples based on Monte Carlo simulations 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 known criteria, and may have superior performance especially in the case of small and contaminated samples.

Romică TRANDAFIR (Technical Univ. of Civil Engineering of Bucharest, Romania)
Estimations of Tsallis and Kaniadakis dynamic cumulative past entropies for power function distribution
joint work with Mihaela PĂUN, Ioana DĂNILĂ, Aura CASARU (Univ. of Bucharest, Romania)

Abstract: In this paper the authors propose estimators of the Tsallis, $\overline{\mathcal{E}}_q^T(t)$, and Kaniadakis, $\overline{\mathcal{E}}_k^K(t)$, past cumulative entropy for the power distribution with two parameters. The MLE and bayesian estimation techniques are used to determine these estimators. Using the Monte Carlo simulation, Baysian estimators for $\overline{\mathcal{E}}_q^T(t)$ and $\overline{\mathcal{E}}_k^K(t)$ are compared under symmetric and asymmetric LINEX loss functions. The authors also derive the highest probability density (HPD) credible intervals for $\overline{\mathcal{E}}_q^T(t)$ and $\overline{\mathcal{E}}_k^K(t)$.

Ciprian TUDOR (Univ. of Lille, France)
Parameter identification for the Hermite Ornstein-Uhlenbeck process
joint work with Obayda ASSAAD (Univ. of Lille, France)

Abstract: By using the analysis on Wiener chaos, we study the behavior of the quadratic variations of the Hermite Ornstein-Uhlenbeck process, which is the solution to the Langevin equation driven by a Hermite process. We apply our results to the identification of the Hurst parameter of the Hermite Ornstein-Uhlenbeck process.

Raluca VERNIC (Ovidius Univ. of Constanta & "Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania)
A neural network approach to prediction from some composite regression models
joint work with Elena PELICAN (Ovidius Univ. of Constanta, Romania)

Abstract: In this paper, we consider two composite regression models obtained by introducing covariates in the lognormal-Pareto and in the lognormal-lognormal composite distributions. We approach the prediction problem from these models by means of regular neural networks, and we discuss the results of this technique on some generated data sets.

Gheorghiță ZBĂGANU ("Gheorghe Mihoc-Caius Iacob" Institute of Mathematical

Statistics and Applied Mathematics, Romania)
Broken stick models

Abstract: A stick of length 1 is divided into smaller sticks according to several algorithms. What happens in the limit? Does the limit exist? It is known that if the breaking points of the stick are iid and uniformly distributed, then the normalized limit is the exponential distribution. It seems that if the breaking points are created according the rule "Hit only the longest stick with uniform distribution", then the normalized limit is the uniform distribution on the interval (0,2). We are not able to prove that. But we prove that if the longest stick is hit in a constant ration, then the limit does not exist - sometimes. We conjecture that it never exists.