21st European Young Statisticians meeting – Book of Abstracts

Published by Faculty of Mathematics, Studentski trg 16, Belgrade

For publisher: Zoran Rakić, dean

Editors: Bojana Milošević, Marko Obradović

Year of publication: 2019

Printed by: Donat Graf d.o.o. Beograd

Number of copies: 76

ISBN: 978-86-7589-135-2

21st European Young Statisticians Meeting

29 July–2 August 2019, Belgrade, Serbia

Book of Abstracts

Eds. Bojana Milošević, Marko Obradović

Preface

This booklet contains basic information about the *21st European Young Statisticians Meeting* (21st EYSM) to be held at Hotel Palace, Belgrade, in the organization of the Faculty of Mathematics, University of Belgrade, from July 28 to August 2 2019.

There are twenty seven European countries participating at the 21st EYSM. The International Organizing Committee (IOC) was responsible for invitation of at most two young scientists from each country whose research interests are in the stochastic range, from pure probability theory to applied statistics. Here the term "young scientist" refers to scientists who are less than thirty years of age or have two to eight years of research experience.

The scientific part of the Conference is organized as follows:

- [-] five eminent scientists from the field of mathematical statistics and probability will hold 60-minutes keynote lectures
- [-] forty eight invited young scientists will hold 20-minutes lectures.

Following the tradition of European Young Statisticians Meetings, there are no parallel sections. The lectures of invited young scientists are divided into eight sessions, six of them having six lectures, and two of them seven lectures each. Sessions were set in such manner that lectures inside one session belong roughly to the same research area, or at least have something in common. However, the sessions were not named since in some cases it was unlikely to find a title that would suit all the lectures within the session.

The topics to be presented include, but are not limited to

- Applied statistics in biology, medicine, ...
- Bayesian inference
- Characterizations of probability distributions
- Extreme and record value theory
- Functional statistics

- Goodness-of-fit testing
- High-dimensional statistics
- Regression models
- Robust estimation
- Spatial statistics
- Stochastic processes
- Survival analysis
- Time series analysis

All invited young scientists have an opportunity to publish a short paper, i.e. extended abstract of their lectures, in the Proccedings of the 21st EYSM. The review process for short papers is organized by the IOC, in the way that the IOC representatives propose reviewers for papers of participants they invited or personally act as a referee.

This booklet, beside all important information about the Conference, contains the scientific program, abstracts of all lectures to be given at the 21st EYSM and the list of participants together with their affiliations and contact information. Abstracts of all contributors are given in order following the schedule of lectures from the scientific program.

More details about the *21st European Young Statisticians Meeting* could be found at the website www.eysm2019.matf.bg.ac.rs

21st European Young Statisticians Meeting

Organized by

Faculty of Mathematics, University of Belgrade

Under the auspices of

Bernoulli Society for Mathematical Statistics and Probability

Ministry of Education, Science and Technological Development of the Republic of Serbia

International Organizing Committee

Apostolos Batsidis, University of Ioannina, Greece

Bettina Porvázsnyik, University of Debrecen, Hungary

Bojana Milošević, University of Belgrade, Serbia

Bruno Ebner, Karlsruher Institut für Technologie (KIT), Germany

David Preinerstorfer, Université libre de Bruxelles, Belgium

Deniz İnan, Marmara University, Turkey

Eduardo García-Portugués, Carlos III University of Madrid, Spain

Johanna Ärje, Tampere University of Technology, Finland

Juan-Juan Cai, Delft University of Technology, Netherlands

Laetitia Teixeira, University of Porto, Portugal

Måns Thulin, Uppsala University, Sweden

Marko Obradović, University of Belgrade, Serbia

Nenad Šuvak, University of Osijek, Croatia

Nina Munkholt Jakobsen, Technical University of Denmark, Denmark

Radim Navrátil, Masaryk University, Czech Republic

Riccardo De Bin, University of Oslo, Norway

Tobias Fissler, Imperial College London, United Kingdom

Wiktor Ejsmont, Mathematical Institute University of Wroclaw, Poland

Local Organizing Committee

Blagoje Ivanović, University of Belgrade, Serbia

Bojana Milošević, University of Belgrade, Serbia

Danijel Subotić, University of Belgrade, Serbia

Marija Minić, University of Belgrade, Serbia

Marko Obradović, University of Belgrade, Serbia

Keynote speakers

Ana Colubi, Justus-Liebig University of Giessen, Germany

Igor Pruñster, Bocconi University, Milan, Italy

M. Dolores Jiménez Gamero, University of Seville, Spain

Pavle Mladenović, University of Belgrade, Serbia

Peter Rousseeuw, KU Leuven, Belgium

Conference structure: keynote lectures, invited lectures.

Conference language: English

Scientific Program

Monday – July 29, 2019

9:05 - 9:30 **Opening ceremony**

9:30 - 11:10 **Session 1**
Chairman: Theodoros Nicoleris

9:30 – 9:55	Alisa Kirichenko *Function estimation on large graphs*
9:55 – 10:20	Jure Vogrinc *Anomalous scaling of the Metropolis Adjusted Langevin Algorithm*
10:20 – 10:45	Federico Camerlenghi *Hierarchical nonparametric processes*
10:45 – 11:10	Alexandre Mösching *Estimating Conditional Distributions under a Stochastic Ordering Constraint*

11:10 - 11:30 **Coffee break**

11:30 - 12:30 **Keynote lecture**

Peter J. Rousseeuw
Department of Mathematics, KU Leuven, Belgium
Detecting Deviating Data Cells

12:30 - 12:40 **Break**

12:40 - 13:30 **Session 1 – continued**
Chairman: Peter J. Rousseeuw

12:40 – 13:05	Nikita Zhivotovskiy *Robust covariance estimation for vectors with bounded kurtosis*
13:05 – 13:30	Arshak Minasyan *On the Convergence and Robustness of Mean Estimation*

13:30 – 15:00 **Lunch**

15:00 - 16:15 **Session 2**
Chairman: Jelena Stanojević

15:00 – 15:25 Stanislav Lohvinenko
Statistical analysis of parameter estimators in the fractional Vasicek model

15:25 – 15:50 Michele Nguyen
From theory to application: a spatio-temporal modelling perspective

15:50 – 16:15 Ashish Kumar
Impact of Stochastic Correlation on Wrong Way risk

16:15 – 16:35 **Coffee break**

16:35 - 17:50 **Session 2 – continued**
Chairman: Fatih Kizilaslan

16:35 – 17:00 Ayça Pamukcu
Risk Prediction in the Presence of Multivariate Repeated Measurements

17:00 – 17:25 Morten Overgaard
The pseudo-observation method

17:25 – 17:50 Blanka Szietl
Controlling Unit-Nonresponse Bias During Within-Household Selection With Optimal Allocation and New Specification of Kish Grid

Tuesday – July 30, 2019

9:05 - 11:10	**Session 3** **Chairman: Alex Karagrigoriou**
	9:05 – 9:30 Jenni Niku *Generalized linear latent variable models with applications*
	9:30 – 9:55 Nuriye Sancar *Liu-type Logistic Estimator based on Particle Swarm Optimization*
	9:55 – 10:20 Vera Djordjilović *Global test for high-dimensional mediation: testing groups of potential mediators*
	10:20 – 10:45 Dario Azzimonti *Bayesian excursion set estimation with Gaussian process models*
	10:45 – 11:10 Una Radojičić *Algorithms for Initialization of Gaussian Mixture Models*
11:10 - 11:30	**Coffee break**
11:30 - 12:30	**Keynote lecture**
	Pavle Mladenović University of Belgrade *Extreme values in samples from stationary sequences and some combinatorial problems*
12:30 - 12:40	**Break**
12:40 - 13:30	**Session 3 – continued** **Chairman: Vladimir Božin**
	12:40 – 13:05 Helene Charlotte Rytgaard *Generalized random forests for survival analysis*
	13:05 – 13:30 Claudia Di Caterina *Fast and efficient selection of high-dimensional graphical models through sparse combination of pairwise scores*

13:30 – 15:00 **Lunch**

15:00 - 16:15 **Session 4**
Chairman: M. Dolores Jiménez Gamero

15:00 – 15:25 Christian Zwatz
Size and Power Properties of Autocorrelation and Heteroskedasticity Robust Tests in Spatial Error Models

15:25 – 15:50 Daniela Correia
GAMLSS with Random Effects: A Tool to Estimate Usual Intake

15:50 – 16:15 Mariusz Kubkowski
Selection consistency of two-step selection method for misspecified binary model

16:15 – 16:35 **Coffee break**

16:35 - 17:50 **Session 4 – continued**
Chairman: Jasmina Đorđević

16:35 – 17:00 Busenur Kızılaslan
Intuitionistic Fuzzy Liu-Type Regression Functions

17:00 – 17:25 Łukasz Rajkowski
A note on the geometry of the MAP partition in some conjugate Normal Bayesian Mixture Models

17:25 – 17:50 Luísa Novais
Order selection in mixtures of linear mixed models

Wednesday – July 31, 2019

9:30 - 11:10 **Session 5**
Chairman: Vlad Stefan Barbu

9:05 – 9:30	Andreas Makrides *Semi-Markov Processes in Reliability: Theory and Applications*
9:30 – 9:55	Céline Cunen *Survival and Competing Risk Models via Gamma Processes*
9:55 – 10:20	Thomas Kuenzer *Spatial PCA for functional random fields*
10:20 – 10:45	Anne van Delft *Frequency domain-based inference of (non-stationary) functional time series*
10:45 – 11:10	Julien Remy *Testing for Principal Component Directions under Weak Identifiability*

11:10 - 11:30 **Coffee break**

11:30 - 12:30 **Keynote lecture**

Ana Colubi
Justus-Liebig-University Giessen
On functional representations to deal with (fuzzy) set-valued data

12:30 - 12:40 **Break**

12:40 - 13:30 **Session 5 – continued**
Chairman: Bojana Milošević

12:40 – 13:05	Javier Álvarez Liébana *A Goodness-of-Fit test for the functional linear model with functional response*
13:05 – 13:30	Daniela Kuruczova *Performance of principal component analysis through conditional expectation on longitudinal data*

13:30 – 15:00 **Lunch**

16:00 - 19:00 **Excursion**

Thursday – August 1, 2019

9:30 - 11:10 **Session 6**
Chairman: Marko Obradović

9:30 – 9:55	Christina Parpoula *Distribution-Free Change-Point Outbreak Detection Control Charts in Biosurveillance*
9:55 – 10:20	Ilia Ragozin *Goodness-of-fit tests for the logistic distribution based on some characterization*
10:20 – 10:45	Slađana Babić *Optimal tests for elliptical symmetry against skew-elliptical alternatives*
10:45 – 11:10	Viktor Skorniakov *On asymptotic normality of certain linear rank statistics*

11:10 - 11:30 **Coffee break**

11:30 - 12:30 **Keynote lecture**

M. Dolores Jiménez Gamero
Departamento de Estadística e Investigación Operativa, Universidad de Sevilla
Goodness-of-fit tests based on the characteristic function

12:30 - 12:40 **Break**

12:40 - 13:30 **Session 6 – continued**
Chairman: Bojana Milošević

12:40 – 13:05	Marija Cuparić *New class of suprem-type exponentiality tests based on V-empirical Laplace transforms and Puri-Rubin characterization*
13:05 – 13:30	Steffen Betsch *Distributional Characterizations for Non-normalized Density Functions and Their Applications*

13:30 - 15:00 **Lunch**

15:00 - 16:15	**Session 7** **Chairman: Jasmina Đorđevic**		
		15:00 – 15:25	Gilles Nisol *Factor Models for Functional Time Series in High Dimensions: Representation Theory and Consistent Estimation*
		15:25 – 15:50	Yolanda Larriba *Order Restricted Inference in Chronobiology*
		15:50 – 16:15	Vít Kubelka *Linear filtering of Gaussian processes in the space of continuous functions*
16:15 – 16:35	**Coffee break**		
16:35 - 17:50	**Session 7 – continued** **Chairman: Lenka Glavaš**		
		16:35 – 17:00	Andrej Gajdoš *Forecasting time series in the light of recent advances in linear mixed modeling and convex optimization*
		17:00 – 17:25	Panagiota Tsamtsakiri *Bayesian model selection for a family of discrete valued time series models*
		17:25 – 17:50	Petra Laketa *On random environment integer-valued autoregressive models – a survey*
20:00 -	**Conference dinner**		

Friday – August 2, 2019

9:30 - 11:10 **Session: 8**
Chairman: Jasmina Đorđevic

9:30 – 9:55	Hrvoje Planinić *Record times of stationary regularly varying time series*
9:55 – 10:20	Niko Leitzén *On blind source separation under martingales: A probability theoretic perspective*
10:20 – 10:45	Kaloyan Vitanov *On decomposable multi-type Bellman-Harris branching process for modeling cancer cell populations with mutations*
10:45 – 11:10	Lívia Leššová *Limit Distribution for Some Iterated Partial Summations*

11:10 - 11:30 **Break**

11:30 - 12:30 **Keynote lecture**

Igor Prünster
Bocconi University, Milan
Bayesian nonparametric models derived from completely random measures

12:45 – 13:15 **Closing ceremony**

13:30 – **Lunch**

Contents

Keynote lectures

Detecting Deviating Data Cells

Peter J. Rousseeuw[1] **and Wannes Van den Bossche**[1]

[1]*Department of Mathematics, KU Leuven, Belgium*

Monday
July 29th
11:30–12:30

Abstract

A multivariate dataset consists of n cases in d dimensions, and is often stored in an n by d data matrix. It is well-known that real data may contain outliers. Depending on the situation, outliers may be (a) undesirable errors which can adversely affect the data analysis, or (b) valuable nuggets of unexpected information. In statistics and data analysis the word outlier usually refers to a row of the data matrix, and the methods to detect such outliers only work when at least half the rows are clean. But often many rows have a few contaminated cell values, which may not be visible by looking at each variable (column) separately. We propose the first method to detect deviating data cells in a multivariate sample which takes the correlations between the variables into account. It has no restriction on the number of clean rows, and can deal with high dimensions. Other advantages are that it provides predicted values of the outlying cells, while imputing missing values at the same time. We illustrate the method on several real data sets, where it uncovers more structure than found by purely columnwise methods or purely rowwise methods. The proposed method can help to diagnose why a certain row is outlying, e.g. in process control. It also serves as an initial step for estimating multivariate location and scatter matrices.

Keywords: Cellwise outlier, Missing values, Multivariate data, Robust estimation, Rowwise outlier.
AMS subject classifications: 62G35.
Acknowledgements: The research of P. Rousseeuw has been supported by projects of Internal Funds KU Leuven. W. Van den Bossche obtained financial support from the EU Horizon 2020 project SCISSOR: Security in trusted SCADA and smart-grids 2015–1017. Thanks go to Jakob Raymaekers for assistance with including this method in the CRAN package *cellWise*.

Bibliography

[1] Agostinelli, C., Leung, A., Yohai, V.J., and Zamar, R.H. (2015). Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination. *Test*, 24, 441–461.

[2] Alqallaf, F., Van Aelst, S., Yohai, V., and Zamar, R.H. (2009). Propagation of outliers in multivariate data. *The Annals of Statistics*, 37, 311–331.

[3] Raymaekers, J., Rousseeuw, P.J., Van den Bossche, W., and Hubert, M. Package 'CellWise', version 2.1.0, CRAN.

[4] Rousseeuw, P.J. and Van den Bossche, W. (2018). Detecting Deviating Data Cells. *Technometrics* 60, 135–145.

Extreme values in samples from stationary sequences and some combinatorial problems

Tuesday
July 30th
11:30–12:30

Pavle Mladenović[1]

[1] *University of Belgrade, Serbia, paja@matf.bg.ac.rs*

Abstract

The basic results related to extreme values in iid sequences of random variables and stationary sequences that satisfy weak dependence conditions will be introduced in the first part of the talk. The second part of the talk will be devoted to extreme values in the incomplete samples from stationary sequences. In the third part of the talk we shall use extreme value theory approach to discuss some combinatorial problems (coupon collector's problem, some extremal properties of random permutations, ...). The possibilities for statistical applications of extreme value theory will be indicated at the end of the talk.

Keywords: Extreme value distributions, Stationary sequences, Weak dependence conditions, Missing observations, Uniform AR(1) processes, Point processes.
AMS subject classifications: 60G70.
Acknowledgements: Research was supported by a grant by Ministry of Education, Science and Technological Development of the Republic of Serbia, Grant No. 174012.

Bibliography

[1] Glavaš, L., Mladenović, P. and Samorodnitsky, G. (2017). Extreme values of the uniform order 1 autoregressive processes and missing observations, *Extremes*, **20**, 671-690.

[2] Glavaš, L. and Mladenović, P. (2018). New limit results related to coupon collector's problem, *Studia Scientiarum Mathematicarum Hungarica*, **55**(1), 115-140.

[3] Glavaš, L., Jocković, J. and Mladenović, P. (2018). Maximum of the sum of consecutive terms in random permutations, *Journal of Statistical Planning and Inference*, **197**, 15-20.

[4] Leadbetter, R. (1974). On extreme values in stationary sequences, *Z. Wahrsch. verw. Gebiete*, **28**, 289-303.

[5] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, Berlin.

[6] Mittal, Y. (1978). Maxima of partial samples in Gaussian sequences. *Annals of Probability*. **6**, 421–432.

[7] Mladenović, P. (2008). Limit distributions for the problem of collecting pairs, *Bernoulli*, **14**(2), 419-439.

[8] Mladenović, P. and Piterbarg, V. (2006). On asymptotic distribution of maxima of complete and incomplete samples from stationary sequences, *Stochastic Processes and their Applications*, **116**, 1977-1991.

On functional representations to deal with (fuzzy) set-valued data

Ana Colubi[1]

[1]*Justus-Liebig-University Giessen, Germany*

Wednesday
July 31st
11:30–12:30

Abstract

Numerous experimental studies involve semi-quantitative expert information, or measured in a non-precise way, which can be modeled with interval (fluctuations, grouped data, etc.) or fuzzy (ratings, opinions, perceptions etc.) data. A general framework to analyze these kinds of inexact data with statistical tools developed for Hilbertian random variables will be presented. The space of nonempty convex and compact (fuzzy) subsets of R^p, has been traditionally used to handle this kind of imprecise data. Mathematically, these elements can be characterized via the support function, which agrees with the usual Minkowski addition, and naturally embeds the considered into a cone of a separable Hilbert space. The support function embedding holds interesting properties, but it lacks of an intuitive interpretation for imprecise data. Moreover, although the Minkowski addition is very natural when $p = 1$, if $p > 1$ the shapes which are obtained when two sets are aggregated are apparently unrelated to the original sets, because it tends to convexity. An alternative and more intuitive functional representation will be introduced in order to circumvent these difficulties. The imprecise data will be modeled by using star-shaped sets on R^p. These sets will be characterized through a center and the corresponding polar coordinates, which have a clear interpretation in terms of location and imprecision, and lead to a natural directionally extension of the Minkowski addition. The structures required for a meaningful statistical analysis from the so-called ontic perspective are introduced, and how to determine the representation in practice is discussed.

Goodness-of-fit tests based on the characteristic function

Thursday
August 1st
11:30–12:30

M. Dolores Jiménez-Gamero[1]

[1]*Departamento de Estadística e Investigación Operativa, Universidad de Sevilla*

Abstract

A classical problem in Statistics is that of checking if the law of the data belongs to a certain parametric family of distributions (the null hypothesis), which is usually called the goodness-of-fit (GoF) problem. A usual strategy to test GoF consists in comparing a nonparametric estimator of a function characterizing the data distribution with a parametric estimator of such function, constructed by assuming that the null hypothesis is true. A common choice for that function is the distribution function, although other choices are possible. In this talk, we review GoF tests based on the characteristic function, putting special emphasis on testing normality.

Keywords: Goodness-of-fit, characteristic function, parametric bootstrap.
AMS subject classifications: 62G09, 62G10.

Bibliography

[1] Baringhaus, L., Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function, *Metrika* 35 339–348.
[2] Epps, T.W., Pulley, L.B. (1983). A test for normality based on the empirical characteristic function, *Biometrika* 70, 723–726.
[3] Henze, N., Jiménez-Gamero, M.D. (2018). A new class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. To appear in *TEST*.
[4] Henze, N., Jiménez-Gamero, M.D., Meintanis, S.G. (2018). A novel characterization of multinormality and corresponding tests of fit, including for GARCH models. To appear in *Econometric Theory*
[5] Jiménez Gamero, M.D. (2014). On the empirical characteristic function process of the residuals in GARCH models and applications, *TEST*, 23, 409–423.
[6] Jiménez Gamero, M.D., Alba Fernández, M.V., Muñoz García, J., Chalco Cano, Y. (2009). Goodness-of-Fit Tests Based on Empirical Characteristic Functions, *Comput. Statist. Data Anal.*, 53, 3957–3971
[7] Jiménez-Gamero, M.D., Kim, H.M. (2015). Fast goodness-of-fit tests based on the characteristic function. *Comput. Statist. Data Anal.* 89, 172–191.
[8] Jiménez-Gamero, M.D., Muñoz-García, J., Pino Mejías R. (2005). Testing goodness of fit for the distribution of errors in multivariate linear models, *Journal of Multivariate Analysis*, 95, 301–322.
[9] Jiménez-Gamero, M.D. Pardo-Fernández. J.C. (2017). Empirical characteristic function tests for GARCH innovation distribution using multipliers. *Journal of Statistical Computation and Simulation*, 87, 2069–2093.
[10] Rivas-Martínez, G.I. Jiménez-Gamero, M.D. (2018). Computationally efficient goodness-of-fit tests for the error distribution in nonparametric regression. *REVSTAT*, 16, 137–166.

Bayesian nonparametric models derived from completely random measures

Igor Prünster[1]

[1] *Department of Decision Sciences and Bocconi Institute for Data Science & Analytics,*
Bocconi University, Milan, Italy

Friday
August 2nd
11:30–12:30

Abstract

The Dirichlet process represents the cornerstone of Bayesian Nonparametrics and is used as main ingredient in a wide variety of models. The many generalizations of the Dirichlet proposed in the literature aim at overcoming some of its limitations and at increasing the models' flexibility. In this talk we provide an overview of a large set of such generalizations by using completely random measures as a unifying concept. All the considered models can be seen as suitable transformations of completely random measures and this allows to highlight interesting distributional structures they share a posteriori in several statistical problems ranging from density estimation and clustering to survival analysis and species sampling. Furthermore, we discuss some natural approaches, which rely on additive, hierarchical and nested structures, to derive dependent versions of Bayesian nonparametric models derived from completely random measures.

Keywords: Bayesian Nonparametrics, Completely random measure, Dirichlet process, Normalized random measure, Posterior distribution.
AMS subject classifications: 62F15; 60G57.

Bibliography

[1] Camerlenghi, F., Dunson, D.B.., Lijoi, A., Prünster, I. and Rodriguez, A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Analysis*, forthcoming.

[2] Camerlenghi, F., Lijoi, A., Orbanz, P. and Prünster, I. (2019). Distribution theory for hierarchical process. *The Annals of Statistics* 47, 67-92.

[3] De Blasi, P., Favaro, S., Lijoi, A., Mena, R., Prünster, I. and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 212-229.

[4] Lijoi, A., Nipoti, B. and Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* 20, 1260-1291.

[5] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C., M'?"?'ller, P., Walker, S.G. Eds.), Cambridge University Press, 80-136.

Abstracts

Function estimation on large graphs

Alisa Kirichenko[1] **and Harry van Zanten**[2]

[1] *Centrum Wiskunde en Informatica (CWI)*
[2] *University of Amsterdam*

Monday
July 29th
9:30–9:55

Abstract

In recent years there has been substantial interest in high-dimensional estimation and prediction problems in the context of relational data. These can in many cases be viewed as high-dimensional or nonparametric regression or classification problems in which the goal is to learn a "smooth" function on a given graph. We present a mathematical framework that allows to study the performance of nonparametric function estimation methods on large graphs and derive the minimax convergence rates within the framework. We consider graphs that satisfy an assumption on their "asymptotic geometry", formulated in terms of the graph Laplacian. We also introduce a Sobolev-type smoothness condition on the target function using the graph Laplacian to quantify smoothness. Finally, we present Bayesian estimation procedures and show how they achieve (asymptotically) optimal regularization.

Keywords: Networks, Bayesian statistics, Asymptotic statistics, Inference on graphs

Anomalous scaling of the Metropolis Adjusted Langevin Algorithm

Jure Vogrinc[1] **and Wilfrid Kendall**[1]

[1] *University of Warwick*

Monday
July 29th
9:55–10:20

Abstract

Metropolis Adjusted Langevin Algorithm (MALA) is a Markov Chain Monte Carlo (MCMC) method that improves on Random walk Metropolis (RWM) algorithm by also using information about the derivatives of the target. For sufficiently smooth targets of product form it can be respectively shown the variance of a single coordinate of proposals in RWM and MALA should optimally scale as d^{-1} and as $d^{-1/3}$ with dimension [1]. Heuristically, the complexity of RWM and MALA grows as d and $d^{1/3}$ with dimension. This leads to the famous guidelines of tuning the

acceptance ratio of RWM and MALA respectively to 0.234 [2] and 0.576 [3]. The work presented will address the open question: to what extent is the requirement of smoothness of the target necessary and how much is it a consequence of the proof? For MALA at least three derivatives of the target are indeed required for the mentioned scaling. I will present a class of generic random targets that are in $\mathcal{C}^2(\mathbb{R})$ but not in $\mathcal{C}^3(\mathbb{R})$, for which the optimal scaling of MALA proposal is almost surely worse than $d^{-1/3}$.

Keywords: Metropolis algorithm, Langevin algorithm, Markov chain Monte Carlo, optimal scaling
AMS subject classifications: 60F05, 60J22, 65C05

Bibliography

[1] Roberts, G.O. and Rosenthal, J.S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science.* 16(4), 351–367.
[2] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability.* 7(1), 110–120.
[3] Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 60(1), 255–268.

Hierarchical nonparametric processes

Monday
29th July
10:20–10:45

Federico Camerlenghi[1]

[1] *University of Milano - Bicocca & Collegio Carlo Alberto*

Abstract

Hierarchical processes are extremely popular Bayesian nonparametric models, which have been successfully applied in linguistics, information retrieval, topic modeling and genetics, among others. They are particularly suited for inducing dependence across different and related samples. We define and investigate suitable classes of hierarchical processes which are useful to face prediction in the context of species or feature models, to estimate random dependent densities or dependent survival functions. More specifically we first focus on hierarchical random probability measures recently introduced and investigated in [2] (see also [3]). Exploiting the results in [2], we are able to devise suitable computational methods to face prediction in the context of species sampling models, when one is provided with multiple populations of animals and wish to predict specific characteristics of additional and unobserved samples. We then move to the context of feature models, which generalize species

sampling models by allowing every observation to belong to more than one species, now called features. We discuss a class of hierarchical priors tailored for such a setting and we introduce all the necessary distributional properties to develop full Bayesian inference (see [1]).

Keywords: Bayesian nonparametrics, prediction, hierarchical processes, species sampling models, feature models
AMS subject classifications: 62F15

Bibliography

[1] Broderick, T., Camerlenghi, F., Favaro, S. and Masoero, L. (2019). Posterior representations of hierarchical completely random measures in trait allocation models. *Manuscript in preparation.*
[2] Camerlenghi, F., Lijoi, A., Orbanz, P. and Prünster, I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.*, 47, 67–92.
[3] Camerlenghi, F., Lijoi, A. and Prünster, I. (2017). Bayesian prediction with multiple–sample information. *J. Multivariate Anal.*, 156, 18–28.

Estimating Conditional Distributions under a Stochastic Ordering Constraint

Alexandre Mösching[1] and Lutz Dümbgen[1]

[1] *University of Bern*

Monday
July 29th
10:45–11:10

Abstract

For a fixed set $\mathcal{X} \subset \mathbb{R}$, consider $n \geq 1$ pairs

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \ \in \ \mathcal{X} \times \mathbb{R},$$

such that, conditional on $\boldsymbol{X} := (X_i)_{i=1}^n$, the responses $Y_1, Y_2, \ldots, Y_n$ are independent with respective distribution functions $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$. That means $\mathbb{P}(Y_i \leq y \,|\, \boldsymbol{X}) = F_{X_i}(y)$, for all $y \in \mathbb{R}$ and $1 \leq i \leq n$. We focus here on a stochastic ordering constraint over the unknown family of distribution functions $(F_x)_{x \in \mathcal{X}}$:

For any fixed $y \in \mathbb{R}$, $F_x(y)$ is decreasing in $x \in \mathcal{X}$. (1)

Such a constraint appears natural on several occasions. For example, an employee's income Y presumably increases with his or her age X. In forecasting, the measured cumulative precipitation amount Y is expected to increase with the numerical predictions X of the same quantity.

The estimation of $(F_x)_{x\in\mathcal{X}}$ under constraint can be done via nonparametric monotone least squares, as discussed in [1], and is often computed with the pool-adjacent-violators algorithm.

One can also consider minimal and maximal β-quantiles of F_x, respectively defined for each $x \in \mathcal{X}$ by $F_x^{-1}(\beta) := \min\{y \in \mathbb{R} : F_x(y) \geq \beta\}$ and $F_x^{-1}(\beta+) := \inf\{y \in \mathbb{R} : F_x(y) > \beta\}$. Then, constraint (1) is equivalent to $F_x^{-1}(\beta)$ and $F_x^{-1}(\beta+)$ being increasing in x, for any fixed $\beta \in (0,1)$. Nonparametric monotone regression quantiles provide estimators for $(Q_x)_{x\in\mathcal{X}}$, with $Q_x(\beta)$ being any β-quantile of F_x, see [2] and [4].

In our manuscript [3], we give a detailed description of the aforementioned estimators. We prove that quantiles of $(\widehat{F}_x)_{x\in\mathcal{X}}$ yield a large family of estimated quantile curves containing the estimators $(\widehat{Q}_x)_{x\in\mathcal{X}}$, but also smoother ones. Finally, under some regularity conditions on $\boldsymbol{X}$ and $(F_x)_{x\in\mathcal{X}}$, we provide rates of convergence for $\widehat{F}_x(y)$ and $\widehat{Q}_x(\beta)$, uniformly over some rectangles.

Keywords: Regression quantiles, stochastic order, uniform consistency.

AMS subject classifications: 62G08, 62G20, 62G30.

Acknowledgements: This work was supported by Swiss National Science Foundation.

Bibliography

[1] El Barmi, H. and Mukerjee, H. (2005). Inferences under a stochastic ordering constraint. *JASA* 100, 252–261.

[2] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* 46, 33–50.

[3] Mösching, A. and Dümbgen, L. (2019). Monotone least squares and isotonic quantiles. *Preprint available at:* https://arxiv.org/abs/1901.02398.

[4] Poiraud-Casanova, S. and Thomas-Agnan, C. (2000). About monotone regression quantiles. *Statist. Probab. Lett.* 48, 101–104.

Robust covariance estimation for vectors with bounded kurtosis

Monday
July 29th
12:40–13:05

Nikita Zhivotovskiy[1]

[1] *This work was prepared while the author was a postdoctoral fellow at the department of Mathematics, Technion I.I.T. and researcher at National University Higher School of Economics. Now visiting researcher at Google.*

Abstract

Let X be a centered random vector taking values in $\mathbb{R}^d$ and let $\Sigma = \mathbb{E}(X \otimes X)$ be its covariance matrix. We show that if X satisfies an $L_4 - L_2$ norm equivalence (sometimes referred to as the bounded kurtosis assumption), there is a covariance estimator $\hat{\Sigma}$ that given a sequence $X_1, ..., X_N$ of independent random vectors distributed according to X exhibits the optimal performance one would expect had X been a gaussian vector. The procedure also improves the current state-of-the-art regarding high probability bounds in the subgaussian case [1] (sharp results were only known in expectation or with constant probability).
In both scenarios the new bound does not depend explicitly on the dimension d, but rather on the effective rank of the covariance matrix Σ.
The talk is based on [2].

Keywords: Robust statistics, Covariance estimation, Matrix Bernstein inequality
AMS subject classifications: Primary 60K35, 60K35; secondary 60K35

Bibliography

[1] Koltchinskii, V. and Lounici K. (2017). Concentration Inequalities and Moment Bounds for Sample Covariance Operators. *Bernoulli.*

[2] Mendelson, S. and Zhivotovskiy, N. (2018). Robust covariance estimation under $L_4 - L_2$ norm equivalence. https://arxiv.org/abs/1809.10462.

On the Convergence and Robustness of Mean Estimation

Arshak Minasyan[1]

[1] *Yerevan State University, Yerevan*

Monday
July 29th
13:05–13:30

Abstract

In [2] the problem of common p-dimensional mean vector estimation of inliers among n independent Gaussian vectors is studied by iteratively using soft-thresholding operator. The presented method is the approximation of the solution of a non-convex optimization problem involving Huber function (see [3]). We simplify this problem and reduce it to Fermat-Weber location problem [1] using Huber function instead of Euclidean distance. Iteratively re-weighted least squares (IRLS) [4] method was modified for minimizing the resulting objective function and the global convergence property was proved given that the starting point is chosen accordingly. We illustrate the robustness of the resulting estimator through numerical experiments and examples, which are nicely consistent with theoretical results.

Keywords: Huber function, Iteratively Re-weighted Least Squares (IRLS), Fermat-Weber problem, Robustness.
AMS subject classifications: 62F35.

Bibliography

[1] J. Blimberg. (1995). The Fermat-Weber location problem revisited. *Mathematical Programming.* 71, 71–76.

[2] O. Collier, A. Dalalyan. (2018). Minimax estimation of a p-dimensional linear functional in sparse Gaussian models and robust estimation of the mean. *arXiv:1712.05495v4.*

[3] P. J. Huber. (1964). Robust estimation of a location parameter. *The annals of mathematical statistics.* 35(1), 73–101.

[4] E. Weiszfeld. (1937). Sur le point pour lequel la somme des distances de n points donnes est minimum. *Tohoku Mathematical Journal.* 43, 355–396.

Statistical analysis of parameter estimators in the fractional Vasicek model

Monday
July 29th
15:00–15:25

Stanislav Lohvinenko[1]

[1] *Taras Shevchenko National University of Kyiv*

Abstract

The fractional Vasicek model, which is described by the following stochastic differential equation

$$X_t = x_0 + \int_0^t (\alpha - \beta X_s)\, ds + \gamma B_t^H, t \geq 0,$$

where B^H is a fractional Brownian motion, is studied. It is assumed that the parameters $x_0 \in \mathbb{R}$, $\gamma > 0$ and $H \in (0,1)$ are known and a problem of estimating α and β is considered.

First, applying the analog of the Girsanov formula for a fractional Brownian motion ([1, Theorem 3]), maximum likelihood estimators (MLEs) for the parameters α and β are obtained for the case $H > 1/2$. The problem is investigated in three variants: *(i)* when β is known, then the MLE for α is derived and it is proved that it is unbiased, strongly consistent and normal ([3, Theorem 3.1]); *(ii)* when α is known, then the MLE for β is described and it is proved that it is strongly consistent and asymptotically normal ([3, Theorem 3.2]); *(iii)* when both parameters are unknown, then the MLEs for α and β are presented and it is proved that they are consistent and asymptotically normal ([3, Theorem 3.3]). Moreover, in the last case, asymptotic distribution of the MLE of vector parameter (α, β) is studied

and its asymptotic normality for $\beta > 0$ is proved ([2, Theorem 4.2]). It is worth noting that the estimators of α and β turned out to be asymptotically independent. Next, the least squares estimators of the unknown parameters are introduced and their strong consistency in the case $H \geq 1/2$ is proved ([4, Theorem 2.1]). Since the discretization and simulation of above estimators are quite difficult, alternative estimators and their discretized versions, which are more useful for practical applications, are presented and their strong consistency for any $H \in (0, 1)$ is proved ([4, Theorems 2.2 and 2.3]).

Keywords: Fractional Brownian motion, fractional Vasicek model, parameter estimation, consistency, asymptotic normality.
AMS subject classifications: 60G22, 62F10, 62F12.
Acknowledgements: I would like to thank Kostiantyn Ralchenko for an invaluable constant support throughout whole my research and Olga Zhuchenko for the theorem about discretized estimators.

Bibliography

[1] Kleptsyna, M. and Le Breton, A. and Roubaud, M.-C. (2000). Parameter estimation and optimal filtering for fractional type stochastic systems. *Stat. Inference Stoch. Process.* 3, 173–182.

[2] Lohvinenko, S. and Ralchenko, K. (2018). Asymptotic distribution of maximum likelihood estimator in fractional Vasicek model. *Teor. Imovir. Mat. Stat.* 99, 134–151 (In Ukrainian).

[3] Lohvinenko, S. and Ralchenko, K. (2017). Maximum likelihood estimation in the fractional Vasicek model. *Lithuanian J. Statist.* 56(1), 77–87.

[4] Lohvinenko, S. and Ralchenko, K. and Zhuchenko, O. (2016). Asymptotic properties of parameter estimators in fractional Vasicek model. *Lithuanian J. Statist.* 55(1), 102–111.

From theory to application: a spatio-temporal modelling perspective

Michele Nguyen[1]

[1]*Malaria Atlas Project, Oxford Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK.*

Monday
July 29th
15:25–15:50

Abstract

Increasingly, data are being collected at higher frequencies and spatial resolutions. To address the inherent complexity, researchers have been developing spatio-temporal models and related methodologies. The most popular ingredient of these

models is the Gaussian random field (RF) with a Matérn covariance. This has been shown to be a solution of a linear fractional stochastic partial differential equation (SPDE) [5].

Recently, there has been renewed interest in viewing the solutions of SPDEs as RFs as well as investigating their probabilistic and statistical properties. Of special mention is the work on ambit fields, a family of non-Gaussian spatio-temporal stochastic integrals, which were introduced in the context of turbulence modelling [1]. We present two fundamental sub-classes of ambit fields: the spatio-temporal Ornstein-Uhlenbeck (STOU) process and the mixed STOU process [3, 4]. By focusing on shape of the integration set and the Lévy basis, we show how these ambit fields are able to model clusters in space-time as well as bridge between short-range and long-range dependence.

In addition to paving the way to more interesting spatio-temporal models, the link between Gaussian Matérn RFs and SPDEs has been used to make inference and prediction with such models computationally feasible [2]. We illustrate this using a study of the seasonality characteristics of malaria cases in Madagascar. A log-linear spatio-temporal regression model is used to predict monthly proportions and fits to von Mises distributions are used to determine useful features such as the start, peak and length of malaria transmission.

Keywords: Spatio-temporal Statistics, random fields, statistical modelling.

AMS subject classifications: 62M30, 62M40, 62M86, 62F10, 62F12, 62F15, 62H11, 62H12.

Acknowledgements: My gratitude goes to Imperial College for my PhD scholarship and the Bill & Melinda Gates Foundation for supporting my current research. I would also like to thank Almut Veraart for the fruitful discussions during my PhD as well as Daniel Weiss, Rosalind Howes and the rest of the Malaria Atlas Project team for their useful feedback on the seasonality work.

Bibliography

[1] Barndorff-Nielsen, O. E., Benth, F. E. and Veraart, A. E. D. (2018). *Ambit Stochastics*, Springer Nature, Switzerland.

[2] Lindgren, F., Rue, H. and Lindstrom J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society, Series B,* 73(4), 423–498.

[3] Nguyen, M. and Veraart, A.E.D. (2017). Spatio-temporal Ornstein-Uhlenbeck processes: theory, simulation and statistical inference. *Scandinavian Journal of Statistics*, 44(1), 46–80.

[4] Nguyen, M. and Veraart, A.E.D. (2018). Bridging between short-range and long-range dependence with mixed spatio-temporal Ornstein-Uhlenbeck processes. *Stochastics,* 90(7), 1023–1052.

[5] Whittle, P. (1954). On stationary processes in the plane. *Biometrika,* 41, 434–449.

Impact of Stochastic Correlation on Wrong Way risk

Ashish Kumar[1], **László Márkus**[1] **and Norbert Hari**[2]

[1] *Eötvös Loránd University, Budapest, Hungary*
[2] *Széchenyi István University, Győr, Hungary,*
[2] *Morgan Stanley, Budapest, Hungary*

Monday
July 29th
15:50–16:15

Abstract

A positive correlation between exposure and counterparty credit risk gives rise to the so called Wrong-Way Risk (WWR). Even after a decade of financial crisis, addressing WWR in a both sound and tractable way remains challenging [1]. Academicians have proposed arbitrage-free set-ups through copula methods but those are computationally expensive and hard to use in practice. Resampling methods are proposed by the industry but they lack in mathematical foundations. This is probably the reason why WWR is not explicitly handled in the Basel III regulatory framework inspite of its acknowledged importance. The purpose of this article is to bridge this gap between the approaches used by academics and industry. To this end, we propose a new method to handle WWR: a stochastic correlation approach in modeling WWR. All the methods proposed post financial crisis more often than not use constant correlation to model the dependency between exposure and counterparty credit risk, i.e. assumes a linear dependency, thus fails to capture the tail dependence. Using a stochastic correlation [3] we move further away from Gaussian copula [2] and can capture the tail risk. This can be achieved by modelling the stochastic correlation as a proper transformation of a diffusion process. For our study we calculate the credit valuation adjustment (CVA) by taking a cross currency swap into account which is prone to wrong way risk because of an additional FX risk other than interest rate risk and credit risk. The performance of our approach is illustrated by a thorough comparison with the case when constant correlation model is used. The results show that even supposing perfect correlation between exposure and credit risk the wrong way risk may be underestimated leading to a wrong calculation of CVA. Given the uncertainty inherent to CVA, the proposed method is believed to provide a promising way to handle WWR in a sound and tractable way.

Keywords: CVA, Wrong Way Risk, Stochastic correlation, Tail dependence, Copulas
AMS subject classifications: 91G40.
Acknowledgements: Széchenyi István University

Bibliography

[1] Damiano BRIGO. and Frdric VRINS. (2018). Disentangling wrong-way risk: pricing credit valuation adjustment via change of measures. *European Journal of Operational Research* Volume 269, Issue 3, 1154-1164.
[2] Roger B. Nelson. (2007). An introduction to Copulas. *Springer Science and Business Media.*
[3] Teng, Long and Ehrhardt, Matthias and Günther, Michael (2016). Modelling stochastic correlation, *Journal of Mathematics in Industry* Volume 6.

Risk Prediction in the Presence of Multivariate Repeated Measurements

Monday
July 29th
16:35–17:00

Ayça Pamukcu[1] and Özgür Asar[1]

[1] *Acıbadem Mehmet Ali Aydınlar University, Biostatistics and Medical Informatics, İstanbul, Turkey*

Abstract

Main objective in many clinical studies is typically to derive accurate prediction of patients' prognosis. For this purpose, generally repeated measurements of a single biomarker are used [1] In the study, we investigate added gains of repeated measurements of additional biomarkers for risk prediction. Joint models for a survival and multivariate longitudinal outcomes under a Bayesian paradigm are used [2]. A real dataset from a randomised clinical trial on primary biliary cirrhosis of the liver patients is analysed. Univariate and bivariate analyses of two biomarkers are considered. Model performances are compared through calibration and discrimination measures. For the specific application, bivariate analysis gave better results.

Keywords: Joint models, Multivariate data, Risk Prediction
AMS subject classifications: 62N99
Acknowledgements: Your acknowledgements.

Bibliography

[1] Hickey, G. L., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology* , 16(1), 117.
[2] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R* . Chapman and Hall/CRC.

The pseudo-observation method

Morten Overgaard[1]

[1]*Department of Public Health, Aarhus University, Bartholins Allé 2, DK-8000 Aarhus C, Denmark*

Monday
July 30th
17:00–17:25

Abstract

This talk will be about the pseudo-observation method. Here, pseudo-observations are jack-knife pseudo-values and are related to the influence of an observation on an estimator. Such pseudo-observations can, under some assumptions, successfully replace incompletely observed outcomes in a regression analysis. This approach is the so-called pseudo-observation method, which is useful in survival analysis where outcomes may be missing due to censoring. The talk will touch upon two challenges in establishing appropriate asymptotic properties. Firstly, unbiasedness of the method relies on the pseudo-observations having, asymptotically, the correct conditional expectation and a challenge is to find out under which circumstances this will be the case. Secondly, the pseudo-observations are clearly correlated and a challenge is to handle this correlation. The correlation of the pseudo-observations turns out to have an impact on the variance of regression parameter estimates which complicates variance estimation.

Controlling Unit-Nonresponse Bias During Within-Household Selection With Optimal Allocation and New Specification of Kish Grid

Blanka Szeitl[1] **and Tamás Rudas**[2]

[1] *Department of Stochastics*
Bolyai Institute, Faculty of Science University of Szeged
[2] *Department of Stochastics*
Bolyai Institute, Faculty of Science University of Szeged

Monday
July 29th
17:25–17:50

Abstract

Several techniques exist to measure and adjust for non-response bias such as propensity models, or post-stratification. All of them can be applied only after the data collection, and require reliable data for the entire population regarding unit non-response patterns however, currently only estimates are available. In this paper,

we demonstrate a new procedure controlling unit non-response during the sampling stage, preceding the actual data collection by combining classical techniques as Neyman's optimal allocation and the Kish grid.
The main finding is that the new sampling algorithm leads to lower SE then the usually applied post stratification.

Keywords: Survey sampling, Unit-Nonresponse, Household selection, Optimal allocation
AMS subject classifications: 62D05

Generalized linear latent variable models with applications

Tuesday
July 30th
9:05–9:30

Jenni Niku[1]

[1] *Department of Mathematics and Statistics, University of Jyväskylä, Finland*

Abstract

In many studies in community ecology, multivariate abundance data are often collected. Such data are characterized by two main features. First, the data are high-dimensional in that the number of species often exceeds the number of sites. Second, the data almost always cannot be suitably transformed to be normally distributed. Instead, the most common types of responses recorded include presence-absence records, overdispersed species counts, biomass, and heavily discretized percent cover data. One promising approach for modelling data described above is generalized linear latent variable models. By extending the standard generalized linear modelling framework to include latent variables, we can account for covariation between species not accounted for by the predictors, species interactions an correlations driven by missing covariates.
We consider the challenges with computationally efficient estimation and inference and introduce methods to overcome these issues, such as the variational approximation and the Laplace approximation method. Using illustrative examples we introduce applications of GLLVMs, eg. constrained or unconstrained ordination, studying species correlation patterns, fourth corner models and making inferences about environmental associations. An R package gllvm for fitting the models is also introduced.

Keywords: Community analysis, ecological data, multivariate data, ordination, variational approximation.
AMS subject classifications: 62H12, 62J12, 62P12.

Bibliography

[1] Hui F.K.C., Warton D.I., Ormerod J.T., Haapaniemi V., Taskinen S. (2017). Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics.* 26:35–43.
[2] Niku J., Brooks W., Herliansyah R., Hui F.K.C., Taskinen S, Warton D.I. (2017) gllvm: R package version 0.1.0.
[3] Niku J., Warton D.I., Hui F.K.C., Taskinen S. (2017). Generalized linear latent variable models for multivariate count and biomass data in ecology. *Journal of Agricultural, Biological, and Environmental Statistics.*22:498–522.

Liu-type Logistic Estimator based on Particle Swarm Optimization

Nuriye Sancar[1] **and Deniz Inan**[2]

[1] *Near East University, Nicosia, Cyprus*
[2] *Marmara University, Istanbul, Turkey*

Tuesday
July 30th
9:30–9:55

Abstract

The logistic regression model is the popular method of regression analysis when the response variable is binary. However, maximum likelihood estimations of the logistic model are severely affected by multicollinearity since this problem causes the inflation of the variance of the maximum likelihood estimator. The Liu-type logistic estimator was introduced as an alternative to the ridge logistic estimator to overcome multicollinearity problem in logistic regression, especially severe multicollinearity problem [2]. In the existing studies for the Liu-type logistic estimator, the pair of shrinkage parameter, k and biasing parameter, d is obtained in two stages [3],[1]. The parameter d is identified after identifying the parameter k, such that the mean squared error of the parameters is minimized. Since it is not feasible to evaluate the mean square error of the parameters numerically, an optimum value of the biasing parameter d is obtained in the terms of parameters. Estimations of different parameters are used to estimate the optimum value of d. Therefore, the pair of the parameters k and d chosen in this way is not optimal. In this study, a new simultaneous approach based on particle swarm optimization is proposed for the estimation of the shrinkage and biasing parameters (k,d) of Liu-type logistic estimator, instead of two stage procedure. Evaluation of the performance of the proposed method is performed by simulation studies and real data sets.

Keywords: Liu-type estimator, Logistic Regression, Particle Swarm Optimization, Multicollinearity
AMS subject classifications: 62J12.

Bibliography

[1] Asar, Y., and Gen, A. (2016). New shrinkage parameters for the Liu-type logistic estimators. *Communications in Statistics-Simulation and Computation* 45, 1094–1103.

[2] Inan, D., and Erdogan, B. E. (2013). Liu-type logistic estimator. *Communications in Statistics-Simulation and Computation* 42, 1578–1586.

[3] Liu, K. (2003). Using Liu type estimator to combat collinearity. *Communications in Statistics: Theory and Methods* 32, 1009-1020.

Global test for high-dimensional mediation: testing groups of potential mediators

Tuesday
July 30th
9:55–10:20

Vera Djordjilović[1]

[1] *Department of Biostatistics, University of Oslo, Norway*

Abstract

Over the last years, we have witnessed an increased interest for causal mediation analysis in genetic epidemiology, genomics, epigenomics, and neuroscience. For instance, researchers in epidemiology investigate whether epigenetic alterations mediate the effect of smoking on lung cancer [2], while in neuroscience, researchers investigate whether brain response measured at thousands of voxels mediates the effect of temperature on reported pain [1].

What these problems have in common is that instead of a single variable M on the path between an exposure X and an outcome Y, there is a high-dimensional vector $\boldsymbol{M}$ (Figure 1). In this work, we address the problem of testing whether $\boldsymbol{M}$ acts as a mediator between X and Y. We propose a global test for mediation, which combines a global test [3] with the intersection-union principle [4]. We derive theoretical properties of our approach and conduct simulation studies which demonstrate that it performs equally well or better than its competitor. We also propose a multiple testing procedure, Screen-Min, that provides asymptotic control

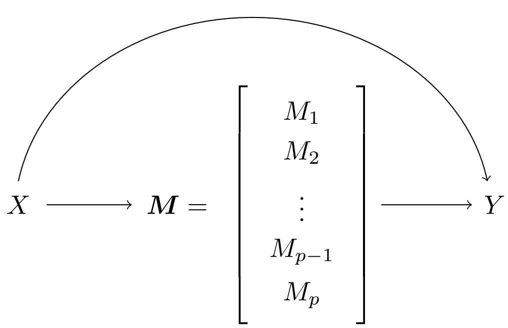

Figure 1: Mediation model with an exposure X, an outcome Y, and a vector $\boldsymbol{M} = (M_1, \ldots, M_p)^\top$ of potential mediators.

of either familywise error rate or false discovery rate when multiple groups of potential mediators are tested simultaneously. We apply our approach to data from a large Norwegian cohort study, where we look at the hypothesis that smoking increases the risk of lung cancer by modifying the level of DNA methylation.

Keywords: global test, mediation analysis, FWER, FDR
AMS subject classifications: 62G10, 62H15 , 62F03, 62P10

Bibliography

[1] Chén OY, Crainiceanu C, Ogburn EL, Caffo BS,Wager TD, Lindquist MA. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics* 2017; 19(2): 121136.
[2] Fasanelli F, Baglietto L, Ponzi E, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nature Communications* 2015; 6: 10192
[3] Goeman JJ, Van De Geer SA, De Kort F, Van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004; 20(1): 9399.
[4] Gleser L. On a Theory of Intersection Union Tests. *Institute of Mathematical Statistics Bulletin* 1973; 2(233): 9.

Bayesian excursion set estimation with Gaussian process models

Dario Azzimonti[1]

[1]*Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland*

Tuesday
July 30th
10:20–10:45

Abstract

We consider the problem of estimating the set of points where an unknown real-valued function is above a certain threshold in the setting where only few function evaluations are available. In a Bayesian framework, we estimate the function with a Gaussian process (GP) regression model and we study the excursion set of the posterior GP distribution. By exploiting random closed set properties, we are able to provide estimates for the excursion set and to quantify their uncertainty. In particular we review sequential design of experiments techniques that reduce the uncertainty on set estimates and we provide a method to explore excursion sets in higher dimensions. We apply the methods to test cases from reliability engineering where the function is an expensive computer experiment and the set of interest represents safe configurations for the system.

Keywords: Gaussian process, excursion set estimation, active learning, experimental design, visualization.
AMS subject classifications: 62M30, 60G15, 62L05, 62P12
Acknowledgements: Dario Azzimonti acknowledges support from the SNSF, grant number 167199. The topic presented is the result of joint works with: D. Ginsbourger, C. Chevalier, J. Bect, and Y. Richet, J. Rohmer and D. Idier.

Bibliography

[1] Azzimonti, D., Bect, J., Chevalier, C., and Ginsbourger, D. (2016). Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA J. Uncertain. Quantif.*, 4(1):850–874.
[2] Azzimonti, D. and Ginsbourger, D. (2018). Estimating orthant probabilities of high-dimensional Gaussian vectors with an application to set estimation. *J. Comput. Graph. Statist.*, 27(2):255–267.
[3] Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2016b). Adaptive Design of Experiments for Conservative Estimation of Excursion Sets. *Under revision.*. arXiv:1611.07256
[4] Azzimonti, D., Ginsbourger, D., Rohmer, J., and Idier, D. (2019). Profile extrema for visualizing and quantifying uncertainties on excursion regions. Application to coastal flooding. *Technometrics*, 0(ja):1–26.
[5] Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.*, 22(3):773–793.
[6] Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

Algoritms for Initialization of Gaussian Mixture Models

Tuesday
July 30th
10:45–11:10

Una Radojičić[1]

[1] *University of J. J. Strossmayer in Osijek - Department of Mathematics*

Abstract

Gaussian mixture models are used for modeling data in cases where the data set is heterogeneous, in a sense that it consists of multiple underlying populations, where each population comes from a certain Gaussian distribution. Gaussian distribution naturally arises as an appropriate distribution for modeling various natural phenomena such as attributes of individuals in a certain population, temperature at certain

area and time of the year, errors that occur because of the imperfect instruments etc. Therefore, data set that consists of observations coming from multiple such populations, is naturally modeled with a Gaussian mixture model. Furthermore, it a Gaussian mixture model is a universal approximator of densities, in the sense that any smooth density can be approximated with any specific nonzero amount of error by a Gaussian mixture model with enough components. The question arises, how to estimate the parameters of such model? While the problem is rather trivial in dimensions one, two and even three, in high-dimensional setting, this problem can be arbitrary hard. A typical approach for parameter estimation in statistical models is a maximization of the likelihood function. The idea of the maximum likelihood estimation is to estimate the model parameters as those for which the observed data is the most likely. The problem in such approach in the case of Gaussian mixture model is that the likelihood function of the model is not convex. Not only that it is not convex, but it can have a huge number of local maxima and saddle points. Therefore, it is easy for numerical local optimization methods, such as for example Expectation Maximization algorithm, especially in high-dimensional problems, to get stuck in one of them, and therefore, fail in attempt of finding a global maximum. The key in solving this problem is finding a good initial approximation to such method. Therefore, we place a focus on algorithms for finding an initial approximation for parameter estimation in Gaussian mixture models.

Keywords: Gaussian Mixture Models, EM Algorithm, Internal Annealing, MCMC, Parameter Estimation
AMS subject classifications: 62G07

Bibliography

[1] Brinda, W. D. (2018). Adaptive Estimation with Gaussian Radial Basis Mixtures. PhD thesis, *Yale University*
[2] Moon, T. K. (1996). The expectation-maximization algorithm. IEEE Signal processing magazine, 13(6), 47-60.

Generalized random forests for survival analysis

Helene Charlotte Rytgaard[1]

[1]*Section of Biostatistics, University of Copenhagen*

Tuesday
July 30th
12:40–13:05

Abstract

Random forests [2] is a machine learning method used for non-parametric prediction in a large variety of problems, including regression analysis, classification and also survival analysis. Generalized random forests (GRFs) [1] are a more recent

extension applicable for inference in more general problems including estimation of average treatment effects. Based on honesty and subsampling, a large sample theory has been developed for GRFs establishing that the estimators are consistent and asymptotically normal.
This talk is concerned with an adaption of the GRF framework to the survival analysis setting where a time-to-event outcome of interest is observed subject to right-censoring. We define our target parameter as the average treatment effect on survival beyond some prespecified time-point, defined as the solution to a martingale estimating equation. In the formulation of treatment effects, potential outcomes already induce a missing data structure to the estimation problem, and the right-censoring of the outcome adds to this missingness in a slightly more complicated way. To customize the partitioning scheme which is central for the GRF algorithm we derive the influence function for our target parameter, written as a functional of the node-specific Kaplan-Meier estimator stratified on the treatment variable. The final estimate of the time-point specific average treatment effect on survival is obtained by using the similarity weights produced by the forest to solve the martingale estimating equation.

Keywords: Survival analysis, random forests, causal inference, treatment effects, influence function.

Bibliography

[1] Athey, Susan and Tibshirani, Julie and Wager, Stefan and others (2019). Generalized random forests. *The Annals of Statistics.* 47.2, 1148–1178.
[2] Breiman, Leo (2001). Random forests. *Machine learning* 45.1, 5–32.

Fast and efficient selection of high-dimensional graphical models through sparse combination of pairwise scores

Tuesday
July 30th
13:05–13:30

Claudia Di Caterina[1], Davide La Vecchia[2] and Davide Ferrari[1]

[1] *Free University of Bozen-Bolzano*
Faculty of Economics and Management
piazza Università 1, 39100 Bozen-Bolzano, Italy.
E-mails: claudia.dicaterina@unibz.it, davide.ferrari2@unibz.it
[2] *University of Geneva*
Geneva School of Economics and Management, Research Center for Statistics
40 Boulevard du Pont-d'Arve, 1211 Geneva 4, Switzerland.
E-mail: davide.lavecchia@unige.ch

Abstract

Undirected graphical models have become important analysis tools in many areas of applied research, including bioinformatics and genetics [1]. For high-dimensional graphs consisting of a large number of potentially meaningful edges, however, both model selection and inference are challenging due to the computational intractability of the full likelihood. We introduce a new methodology for simultaneous model selection and parameter estimation in large graphical models. Within a pairwise likelihood framework [2], our proposal is to build estimating equations by including only the most informative pairwise likelihood scores selected from a great variety of candidates. Selection is carried out by minimizing an approximate distance between the composite and the full likelihood score functions, subject to a ℓ_1-constraint. Differently from other approaches based on penalized likelihoods with penalties for model parameters [3, 4, 5], here penalization occurs on coefficients for the pairwise likelihood scores. This strategy discourages the construction of overly complex likelihoods, thus reducing the computing burden while retaining unbiasedness of the resulting parameter estimator. The performance of our method is illustrated through numerical simulations and real data examples.

Keywords: composite likelihood estimation, lasso penalty, model selection, regularization.
AMS subject classifications: 62H12, 62M40, 62P10.

Bibliography

[1] Jordan, M. I. (2004). Graphical models. *Statistical Science*, 19, 140–155.
[2] Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80, 221–239.
[3] Meinshausen, N., and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34, 1436–1462.
[4] Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics*, 38, 1287–1319.
[5] Yang, E., Allen, G., Liu, Z., and Ravikumar, P. K. (2012). Graphical models via generalized linear models. *Advances in Neural Information Processing Systems*, 25, 1358–1366.

Size and Power Properties of Autocorrelation and Heteroskedasticity Robust Tests in Spatial Error Models

Tuesday
July 30th
15:00–15:25

Christian Zwatz[1]

[1] *University of Vienna*

Abstract

A typical approach for testing linear hypotheses on the regression parameters in regression models with autocorrelated and/or heteroskedastic disturbances is to modify the conventional F-test statistic by using a heteroskedasticity and autocorrelation consistent (HAC) estimator for the covariance matrix. These are nonparametric estimators designed to take the heteroskedasticity and autocorrelation in the data into account. HAC estimators for models where heteroskedasticity and autocorrelation is due to spatial interdependence among the observational units have been, e.g., proposed in [1] and [2].
We consider heteroskedasticity and autocorrelation robust testing in spatial error models, i.e. models where the disturbances follow a spatial autoregressive or spatial moving average process. It is well known that tests based on HAC estimators in the case of time series regression models suffer from substantial size and power problems. Based on a general theory developed in [3], [4] and [5], we show that similar problems also occur in the spatial error model. In particular, we give conditions under which the size of the resulting test is in fact one. We also give conditions under which the size of the test can be controlled by an appropriate choice of critical value.

Keywords: Size distortion, power deficiency, HAC, spatial models
AMS subject classifications: 62F03, 62F05, 62J05, 62M30, 91B72

Bibliography

[1] Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.
[2] Pinkse, J., Slade, M. E. and Brett, C. (2002). Spatial Price Competition: A Semiparametric Approach. *Econometrica* 70, 1111–1153
[3] Pötscher, B. M. and Preinerstorfer, D. (2018). Controlling the Size of Autocorrelation Robust Tests. *Journal of Econometrics* 207, 406–431.
[4] Pötscher, B. M. and Preinerstorfer, D. (2017). Further Results on Size and Power of Heteroskedasticity and Autocorrelation Robust Tests, with an Application to Trend Testing. arXiv: 1708.08688
[5] Preinerstorfer, D. and Pötscher, B. M. (2016). On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests. *Econometric Theory* 32, 261–358.

GAMLSS with Random Effects: A Tool to Estimate Usual Intake

Daniela Correia[1,2,3], **Milton Severo**[2,3] **and Óscar Felgueiras**[1,4]

[1]*Centre of Mathematics of the University of Porto, Portugal*
[2]*EPIUnit, Institute of Public Health, University of Porto, Portugal*
[3]*Department of Public Health and Forensic Sciences, and Medical Education, Unit of Epidemiology, Faculty of Medicine, University of Porto, Portugal*
[4]*Mathematics Department, Faculty of Sciences of the University of Porto, Portugal*

Wednesday
July 30th
15:25–15:50

Abstract

Generalized Additive Models for Location, Scale and Shape (GAMLSS) are an extension of Generalized Additive Models (GAM) that allow parametric distributions from outside the exponential family for the response variable as well as modeling location, scale and shape parameters as linear or smooth functions of explanatory variables. It is a versatile yet simple method that allows regression predictors to be placed on any parameter of a potentially complex response distribution. While GAMLSS was originally designed for independent observations, the fact that smoothing can be represented as a random effect model facilitates the idea of introducing random effects within GAMLSS.
The usual intake of a food or nutrient of an individual is defined as the long-run average daily intake focusing on the long-term patterns rather than consumption on any given day. It is extremely relevant to some chronic phenomena such as inadequate dietary intake or long-term exposure to food contaminants. Its direct observation is unpractical since it would require participants to complete dietary intake questionnaires over a long period of time such as a year, depending on the foods/nutrient to be measured. Hence, each participant is asked to report dietary intakes of a few randomly selected days, so that usual intake can be estimated at the population level, as a repeated measurements distribution. However, it is necessary to consider that individuals day-to-day variability in consumption produces large within-person variance, heterogeneous across subjects. For that reason, the population usual intake distribution is the result distribution of average daily intakes recorded after removing within-person variability. Therefore, GAMLSS with random effects can be used to estimate the distribution of usual intake of foods/nutrients in a population. Thus, with this work we aim to explain how random effects models can be used within GAMLSS and explain the advantages and disadvantages of such modeling. Also, we use these models to estimate usual intake distributions and compare it to the results obtained using the standard approaches to estimate usual intake. We use data from the National Food, Nutrition, and Physical Activity Survey of the Portuguese General Population, IAN-AF 2015-2016.

Keywords: gamlss, random effects, usual intake.

AMS subject classifications: 62P10, 92D30.
Acknowledgements: Data used throughout the present study are retrieved from the IAN-AF databank, developed in the context of the National, Food, Nutrition and Physical Activity Survey funded by the EEAGRANTS programme-initiatives in Public Health (EEAGRANTS PT06_00088SI3). The project is coordinated by the University of Porto. The author is solely responsible for the contents of the document. The opinions expressed do not represent the opinions of the Consortium and the Consortium is not responsible for any use that might be made of the information included.

Selection consistency of two-step selection method for misspecified binary model

Tuesday
July 30th
15:25–15:50

Mariusz Kubkowski[12] **and Jan Mielniczuk**[12]

[1] *Warsaw University of Technology*
[2] *Institute of Computer Science Polish Academy of Sciences*

Abstract

We consider two-stage selection method of predictors $\mathbf{X} \in \mathbb{R}^p$ when the underlying random binary regression model:

$$\mathbb{P}(Y = 1|\mathbf{X} = \mathbf{x}) = q(\mathbf{x}) \tag{2}$$

is misspecified. We discuss a problem of finding consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^*$ minimizes risk function:

$$R(\mathbf{b}) = \mathbb{E}\rho(\mathbf{b}^T\mathbf{X}, Y)$$

for $\mathbf{b} \in \mathbb{R}^p$ and $\rho\colon\ \mathbb{R}\times\{0,1\} \to \mathbb{R}$ is convex function. We call model (2) misspecified, when $q(\mathbf{x}) = q(\boldsymbol{\beta}^T\mathbf{x})$ and corresponding minus log-likelihood is not equal ρ. In this case an aim of selection is to find $\hat{\boldsymbol{\beta}}$ which recovers the support of $\boldsymbol{\beta}^*$ with high probability. The proposed procedure consists of screening and ordering predictors by Lasso and then selecting a subset of predictors which minimizes Generalized Information Criterion on the nested family pertaining to them.
In the contribution we discuss sufficient conditions in [1] on the parameters of the method and distribution of (X, Y) under which the above procedure is consistent.
In numerical experiments we discuss performance of several modifications of the above procedure, in particular its net version when the nested family $\mathcal{M}$ is replaced by the sum of such families constructed for a net of λs.

Keywords: misspecification, binary regression, consistent selection, subgaussianity, Generalized Information Criterion
AMS subject classifications: Primary 62J05, 62J07, secondary 90C25

Bibliography

[1] Kubkowski, M. and Mielniczuk, J. (2018). Selection consistency of two-step selection method for misspecified logistic regression, *in preparation.*

Intuitionistic Fuzzy Liu-Type Regression Functions

Busenur Kızılaslan[1], Erol Egrioglu[2] and Atif Ahmet Evren[3]

[1] *Department of Statistics, Marmara University, Istanbul, Turkey*
[2] *Department of Statistics, Giresun University, Giresun, Turkey*
[3] *Department of Statistics, Yildiz Technical University, Istanbul, Turkey*

Tuesday
July 30th
16:35–17:00

Abstract

Developing technology shows how useful fuzzy inference systems in lots of applications. Fuzzy functions approach [3] which is one of the important fuzzy inference system for time series forecasting. In fuzzy functions approach, the membership values and their non-linear transformations are used together with original input variables to increase the prediction power. However, multicollinearity problem can be occured because of using these correlated variables. Purpose of the paper is to propose a new fuzzy forecasting method with intuitionistic fuzzy sets [1] which has addition information known as hesitation degree. In this case, both intuitionistic fuzzy sets and their non-linear transformations is used to increase the prediction power. Liu-Type estimator [2] is preferred to obtain intuitionistic fuzzy functions without exposed to multicolinearity problem. To demonstrate the performances of proposed method, some real world time series data are used and the results have shown that the effectiveness of the proposed method in conrast to other methods.

Keywords: forecasting, intuitionistic fuzzy sets, liu-type regression, multicollinearity, ridge regression
AMS subject classifications: 62J07, 62M10, 94D05

Bibliography

[1] Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1), 87–96.
[2] Liu, K. (2003). Using Liu-Type Estimator to Combat Collinearity. *Communications in Statistics*, 32(5), 1009–1020.
[3] Turksen, I. B. (2008). Fuzzy function with LSE. *Applied Soft Computing*, 8, 1178–1188.

A note on the geometry of the MAP partition in some conjugate Normal Bayesian Mixture Models

Tuesday
July 30th
17:00–17:25

Łukasz Rajkowski[1] and John Noble[1]

[1]*Faculty of Mathematics, Informatics and Mechanics, University of Warsaw*

Abstract

Bayesian Mixture Models are increasingly popular among researchers for clustering problems. I will present the results of our research regarding the geometry of the maximal a posteriori (MAP) partition in the Bayesian Mixture Model where the component distribution is multivariate Normal with Normal-inverse-Wishart prior on the component mean and covariance ([2]). We prove that in this case the clusters in any MAP partition are quadratically separable. Basically this means that every two clusters are separated by a quadratic surface. In connection with results of [1], where the linear separability of clusters in the Bayesian Mixture Model with a fixed component covariance matrix was proved, it gives a nice Bayesian analogue of the geometric properties of Fisher Discriminant Analysis (LDA and QDA). Moreover, I will show how the first order approximation of the posterior probability gives an elegant formula which may serve as a score function in choosing the optimal clustering (for example) hierarchical clustering methods. This is the topic of our ongoing work.

Keywords: Bayesian Mixture Models, Maximal a Posteriori Partition
AMS subject classifications: 62F15

Bibliography

[1] Rajkowski, Ł. (2018). Analysis of the maximal a posteriori partition in the Gaussian Dirichlet Process Mixture Model., *Bayesian Analysis*, advance publication, 30 July 2018. doi:10.1214/18-BA1114.
[2] Rajkowski, Ł. and Noble, J. (2019). A note on the geometry of the MAP partition in some Normal Bayesian Mixture Models., *arXiv preprint* arXiv:1902.01141

Order selection in mixtures of linear mixed models

Tuesday
July 30th
17:25–17:50

Luísa Novais[1] and Susana Faria[2]

[1,2]*Department of Mathematics and Applications, University of Minho, Portugal*

Abstract

Finite mixture models are one of the most widely used methods for modelling data that arise from a heterogeneous population, given that for these cases the estimation of a single linear model is not sufficient. In regression analysis, it has been a popular practice to model unobserved population heterogeneity through finite mixtures of regression models. Within the family of mixtures of regression models, finite mixtures of linear mixed models have also been applied in different areas of application since they not only take into account the correlations between observations from the same individual, but they also conveniently model unobserved heterogeneity between individuals at the same time.

One of the main difficulties in mixture models arises in the selection of the correct number of components for each case, an important and difficult research problem that has not yet been resolved. In this work we address the problem of determining the number of components for mixtures of linear mixed models by investigating the performance of various model selection methods, namely different information criteria and the likelihood ratio test.

In order to evaluate the methodologies developed we carry out a simulation study, in particular using resampling methodologies, and we apply these methodologies to a real data set. The results demonstrate that the criteria *HQIC*, *AIC4* and *aBIC*, as well as the likelihood ratio test, are the best options to estimate the number of components for mixtures of linear mixed models.

Keywords: Mixture models, Model selection, Information criteria, Likelihood ratio test, Bootstrap

AMS subject classifications: 62J99

Acknowledgements: This research was financed by FCT - Fundação para a Ciência e a Tecnologia, through the PhD scholarship with reference SFRH/BD/139121/2018. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684.

Bibliography

[1] Bai, X., Chen, K. and Yao, W. (2016). Mixture of linear mixed models using multivariate t distribution. *Journal of Statistical Computation and Simulation*, 86, 771–787.

[2] Depraetere, N. and Vandebroek, M. (2014). Order selection in finite mixtures of linear regressions. *Statistical Papers*, 55, 871–911.

[3] Li, M., Xiang, S. and Yao, W. (2016). Robust estimation of the number of components for mixtures of linear regression models. *Computational Statistics*, 31, 1539–1555.

[4] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, John Wiley & Sons.

[5] Young, D. and Hunter, D. (2015). Random effects regression mixtures for analyzing infant habituation. *Journal of Applied Statistics*, 42, 1421–1441.

Semi-Markov Processes in Reliability: Theory and Applications

Wednesday
July 31st
9:05–9:30

Andreas Makrides[1], **Alex Karagrigoriou**[2] **and Vlad Stefan Barbu**[1]

[1] *University of Rouen*
[2] *University of the Aegean*

Abstract

This work deals with multi state systems that we model by means of semi-Markov processes. The main characteristic of this work is that the sojourn times in a given state which are seen to be independent not identically distributed random variables are assumed to belong to two different general classes of distributions. The first class of distributions is closed under maxima and contains several distributions, like the Bernoulli distribution, the power function distribution and the extreme value Type I distribution. The second class of distributions is closed under minima and includes the exponential, the Weibull, the Pareto, the Rayleigh and the Erlang truncated exponential distribution [1]. For the above setting we obtain maximum likelihood estimators of the parameters of interest and investigate their asymptotic properties. Furthermore, plug-in type estimators are furnished for various reliability indices related to the system under study.
Conclusively, our main objective is the proposal of parsimonious modeling for multi-state systems, considering also a semi-Markov framework. Thus, we introduce a useful and powerful tool with a reduced number of parameters, which can be of great importance from a practical point of view.

Keywords: multi-state system, reliability theory, survival analysis, semi-Markov processes, parameter estimation.
AMS subject classifications: 60J27, 60K15, 60K20, 62N02, 62N05, 90B25.
Acknowledgements: This work has been carried out during the post-doc appointment of the first author at the Laboratoire de Mathématiques Raphaël Salem, University of Rouen, France.

Bibliography

[1] Balasubramanian, K., Beg, M. I. and Bapat, R. B. (1991). On families of distributions closed under extrema, *Sankhya: The Indian Journal of Statistics A*, 53, 375-388.
[2] Barbu, V. S., Brard, C., Cellier, D., Sautreuil, M. and Vergne N. (2017). Parametric estimation of semi-Markov chains (submitted).
[3] Barbu, V. S., Karagrigoriou, A. and Makrides, A. (2017). Semi-Markov modelling for multi-state systems. *Meth. & Comput. Appl. Prob.* 19, 1011–1028.
[4] Limnios, N. and Oprişsan, G. (2001). *Semi-Markov Processes and Reliability*, Birkhäuser, Boston.

Survival and Competing Risk Models via Gamma Processes

Céline Cunen[1] **and Nils Lid Hjort**[1]

[1] *University of Oslo, Department of Mathematics*

Wednesday
July 31st
9:30–9:55

Abstract

We use gamma processes as a vehicle to construct biologically plausible models for survival and competing risk models. Our methods include both time-to-threshold perspectives, where an event takes place when the underlying process reaches such a threshold c (see for instance [1]), and sizes of shocks, where the event happens when a shock is above a different threshold d. These considerations lead to fruitful and flexible classes of survival distributions, also for the regression case with covariate information for the individuals under study. In addition, the time-to-threshold models are readily extended to the competing risk case, where one wishes to study different causes leading to the same event. Our models and methods are applied to real data.

Keywords: survival analysis, competing risk, gamma processes.
AMS subject classifications: 62N99, 62P10

Bibliography

[1] Lee, M.-L. T. & Whitmore, G. A. (2006). Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary. *Statistical Science* 21, 501–513.

Spatial PCA for functional random fields

Thomas Kuenzer[1], **Siegfried Hörmann**[1] **and Piotr Kokoszka**[2]

[1] *Institute of Statistics, Graz University of Technology, Austria*
[2] *Statistics Department, Colorado State University, Fort Collins, CO, USA*

Wednesday
July 31st
9:55–10:20

Abstract

With the proliferation of remote sensing technology, functional spatial data is a subject of growing importance in statistics. The applications of such high-dimensional functional data on spatial grids range from climate data to hyperspectral imaging of

the earth's surface. We propose a novel concept of functional principal components (FPCs) for spatial data. Ordinary functional principal component analysis ignores the information contained in the spatial dependence structure of a sample. By making use of this information, our method of spatial functional principal components (SFPC) provides a very efficient tool for dimension reduction.
We consider second-order stationary functional data on a spatial grid $\mathbb{Z}^r$ for some $r \geq 1$, where in each grid point $\mathbf{t} \in \mathbb{Z}^r$, a random element $X_{\mathbf{t}} \in L^2([0,1])$ can be observed. Under mild assumptions on the functional random field $(X_{\mathbf{t}})_{\mathbf{t} \in \mathbb{Z}^r}$, we show that SFPCs provide both a quantitative and a qualitative improvement to the dimension reduction technique provided by the ordinary FPCs. That is, a given number of SFPCs explains a larger portion of the variance than the respective ordinary FPCs, while the SFPC scores exhibit a cross-covariance of zero at all spatial lags. The estimators of the SFPCs and their scores are shown to be consistent under the framework of increasing domain asymptotics. In empirical studies on real and simulated data, we show that SPFCs significantly outperform ordinary FPCs in terms of the mean squared error of the reconstructed data. We also briefly explore a possible statistical application of this method in the context of normality testing.

Keywords: Dimension reduction, Functional data analysis, Frequency domain analysis, Functional spatial data, Principal components.
AMS subject classifications: 62M30, 62H25 (primary) 62M15, 62G20, 62G05 (secondary)

Frequency domain-based inference of (non-stationary) functional time series

Wednesday
July 31st
10:20–10:45

Anne van Delft[1]

[1]*Ruhr University Bochum, Germany*

Abstract

In the analysis of functional time series the objects of interest are ordered collection of random variables $\{X_t, t \in \mathbb{Z}\}$ where each X_t takes values in some function space, often assumed to be L^2. The setting is therefore high-dimensional where the intrinsic variation is inherently larger than the observation length. In order to extract information on the characteristics of such series, frequency domain-based inference techniques provide powerful tools: not only for optimal dimension reduction of stationary functional time series but also to model non-stationary functional time series as well as for hypothesis testing. We review some of the existing techniques and introduce some new results.

Keywords: Functional data analysis, locally stationary processes, spectral analysis.
AMS subject classifications: Primary: 62HG99, 62M15, Secondary 62M10, 62H15.

Testing for Principal Component Directions under Weak Identifiability

Davy Paindaveine[1], **Julien Remy**[1] **and Thomas Verdebout**[1]

[1]*Universit Libre de Bruxelles*

Wednesday
July 31st
10:45–11:10

Abstract

We consider the problem of testing, on the basis of a p-variate Gaussian random sample, the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$, where $\boldsymbol{\theta}_1$ is the "first" eigenvector of the underlying covariance matrix and $\boldsymbol{\theta}_1^0$ is a fixed unit p-vector. In the classical setup where eigenvalues $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_p$ are fixed, the Anderson likelihood ratio test [1] and the Hallin, Paindaveine, Verdebout Le Cam optimal test [2] for this problem are asymptotically equivalent under the null, hence also under sequences of contiguous alternatives. We show that this equivalence does not survive asymptotic scenarios where $\lambda_{n1} - \lambda_{n2} = o(r_n)$ with $r_n = O(1/\sqrt{n})$. For such scenarios, the Le Cam optimal test still asymptotically meets the nominal level constraint, whereas the LRT becomes extremely liberal. Consequently, the former test should be favored over the latter one whenever the two largest sample eigenvalues are close to each other. By relying on the Le Cam theory of asymptotic experiments, we study in the aforementioned asymptotic scenarios the non-null and optimality properties of the Le Cam optimal test and show that the null robustness of this test is not obtained at the expense of efficiency. Our asymptotic investigation is extensive in the sense that it allows r_n to converge to zero at an arbitrary rate. To make our results as striking as possible, we not only restrict to the multinormal case but also to single-spiked spectra of the form $\lambda_{n1} > \lambda_{n2} = \ldots = \lambda_{np}$.

Keywords: Testing for directions, Weak identifiability, Le Cam, LAN, PCA
AMS subject classifications: 62H11

Bibliography

[1] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics* 34, 122–148.
[2] Hallin, M., Paindaveine, D. and Verdebout, Th. (2010). Optimal rank-based testing for principal components. *Annals of Statistics*, 38, 3245–3299.
[3] Paindaveine, D., Remy, J. and Verdebout, Th. (2019). Testing for principal component directions under weak identifiability. *Annals of Statistics*, to appear.

A Goodness-of-Fit test for the functional linear model with functional response

Wednesday July 31st 12:40–13:05

Javier Álvarez-Liébana[1], E. García-Portugués[2], G. Álvarez-Pérez[3], W. González-Manteiga[4] and M. Febrero-Bande[4]

[1] *Contract Professor at Department of Statistics and O.R., University of Oviedo (Spain)*
[2] *Assistant Professor at the Department of Statistics, Carlos III University of Madrid (Spain)*
[3] *PhD student at Department of Physics, University of Oviedo (Spain)*
[4] *Full Professor at Department of Statistics and O.R., University of Santiago de Compostela (Spain)*

Abstract

The increasing availability of data for continuous processes has recently boosted the field of Functional Data Analysis (FDA). There is a large body of situations in which it is desirable to measure the relation between functional random variables $\mathcal{Y}$ and $\mathcal{X}$, by means of a linear regression $\mathcal{Y} = m(\mathcal{X}) + \mathcal{E}$, named as Functional Linear Model with Functional Response (FLMFR), where $\mathcal{E}$ is a random error function, being m an unknown operator, admitting an integral representation in terms of a bivariate kernel β. Several authors have contributed to the Goodness-of-Fit (GoF) framework for scalar and multivariate regression models. Overcoming the poor empirical power of smoothing-based tests and tests based on the integrated regression function, a novel statistic, in the high-dimensional context, was proposed by [2], in terms of a residual marked empirical process based on projections. In contrast, the development of GoF test approaches in the FLMFR context is currently one of the challenges in FDA. Up to our knowledge, only proposals in [1] (just a residual diagnostic), [4, 5] (just tested the simple hypothesis) and [6] (only for concurrent models) can be found. In this work we will extend the GoF test in [3] for the FLMSR (FLM with Scalar Response) model (based on the proposal in [2]): we will present an easily computable Cramér-von-Mises statistic for the null composite hypothesis, being characterized in terms of an infinite–dimensional process by its two-sides projected version into functional directions, and the empirical estimator of the integrated regression function depending on empirical marked processes. Since β should be estimated, we also provide a novel estimator based on Lasso regression. The calibration of the test is performed by a wild bootstrap, applied to the functional residuals.

Keywords: CvM statistic, Functional linear model, Functional response, Goodness-of-Fit, Wild bootstrap.
AMS subject classifications: 62J02; 62M05; 62M10; 62M20
Acknowledgements: The work of J. Álvarez-Liébana was supported in part by project MTM2015-71839-P (co-funded by Feder funds), of the DGI, MINECO, Spain.

Bibliography

[1] Chiou, J. M. and Müller, H. G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* 51, 4849–4863.
[2] Escanciano, J. C. A consistent diagnostic test for regression models using projections. *Econometric Theory* 22, 1030–1051.
[3] García–Portugués, E., González–Manteiga, W. and Febrero–Bande, M. (2014). A Goodness-of-Fit Test for the Functional Linear Model with Scalar Response. *J. Comp. Graph. Stat.* 23, 761–778.
[4] Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* 36, 207–222.
[5] Patilea, V, Sánchez–Sellero, C. and Saumard, M. (2016). Testing the Predictor Effect on a Functional Response . *J. Amer. Statist. Assoc.* 111, 1684–1695.
[6] Wang, H., Zhong, P. S., Cui, Y. and Li, Y. (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *J. R. Statist. Soc. B* 60, 343–364.

Performance of principal component analysis through conditional expectation on longitudinal data

Daniela Kuruczová[1]

[1]*Department of Mathematics and Statistics, Faculty of Science, Masaryk University*

Wednesday
July 31st
13:05–13:30

Abstract

Functional data analysis is a relatively novel approach in multivariate statistics which extends the concept of a random variable to the infinite dimensional setting. While the measurements of a functional random variable still have finite dimension, it is assumed, that the observed variable is in its nature infinitely dimensional. Natural examples of functional variables are time-dependent functions, but the notion can be extended to many other concepts, such as surfaces [1].
Longitudinal studies are a natural example of data suitable for the functional approach. Multiple variables are measured at several discrete points over time with the intention of estimating the underlying continuous function. The usual technique for analysing functional data is principal component analysis, which may be unsuitable in case of longitudinal data due to sparseness. To address this issue, a new method was developed – principal component analysis through conditional expectation (PACE). This method heavily relies on a non-parametric estimate of the mean and covariance function [2].

Via simulation study, we examined how the quality of these estimates affects the performance of the principal component analysis through conditional expectation, focusing on the choice of the smoothing parameter and its extremes such as over- and undersmoothing. Then, using simulations as well as an actual dataset, we assessed the performance of this method as an option for dealing with missing data in longitudinal studies, addressing specific challenges of longitudinal data, such as drop-out effect.

Keywords: functional data, principal component analysis, PACE, bandwidth selection, sparse data
AMS subject classifications: 62H99

Bibliography

[1] Kokoszka, P., & Reimherr, M. (2017). *Introduction to functional data analysis.* CRC Press.
[2] Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470), 577-590.

Distribution-Free Change-Point Outbreak Detection Control Charts in Biosurveillance

Thursday
August 1st
9:30–9:55

Christina Parpoula[1] and Alex Karagrigoriou[1]

Lab of Statistics and Data Analysis,
Department of Statistics and Actuarial-Financial Mathematics,
University of the Aegean, Karlovasi, Samos 83200, Greece

Abstract

Epidemiological surveillance is a specific sector of Biosurveillance related solely to the human population and constitutes a dynamic scientific activity which progresses and requires systematic monitoring of developments in the field of health sciences and (bio)statistics. Epidemiological surveillance systems present various challenges regarding the source and quality of data, the statistical quality control, the monitoring (follow-up), the evaluation of statistical techniques used to detect outbreaks, anomalies and outliers in surveillance data, and extreme timeliness of detection [3]. Statistical process control (SPC) techniques have a long history of application to problems in health care monitoring and public health surveillance, and several proposed approaches for detecting outbreaks of infectious diseases are directly inspired by, or related to, methods of SPC [2]. This paper deals with the development of

univariate distribution-free SPC charting methods based on change-point analysis [1], and discusses some of the statistical issues involved in the selection, construction and evaluation of such control charts for Phase I retrospective epidemiological surveillance purposes. Empirical study provides some guidelines for the implementation and effective use of Phase I univariate SPC charting methods for the very early and accurate outbreak detection. The interpretation of such SPC charts, from both a statistical and an epidemiological perspective, facilitates the better understanding of process variability in an epidemiological surveillance system and highlights important issues either for a health care or SPC practitioner.

Keywords: change-point analysis, statistical process control, distribution-free, control charts, outbreak detection.
AMS subject classifications: 62-07, 62-09, 62G05, 62G08, 62G10, 62P10.
Acknowledgements: The work was carried out at the Lab of Statistics and Data Analysis of the University of the Aegean. The authors would like to thank the Department of Epidemiological Surveillance and Intervention of the Hellenic Centre for Disease Control and Prevention (HCDCP) for providing the influenza-like illness (ILI) rate data, collected weekly through the sentinel surveillance system.

Bibliography

[1] Capizzi, G. and Masarotto, G. (2013). Phase I Distribution-Free Analysis of Univariate Data. *Journal of Quality Technology*, 45, 273–284.
[2] Montgomery, D.C. (2013). *Introduction to Statistical Quality Control*, John Wiley and Sons, New Jersey.
[3] Parpoula, C., Karagrigoriou, A. and Lambrou, A. (2017). Epidemic Intelligence Statistical Modelling for Biosurveillance, *In J. Blömer et al. (Eds.): MACIS 2017, Lecture Notes in Computer Science (LNCS)*, LNCS volume 10693, pp. 349-363, Springer International Publishing AG.

Goodness-of-fit tests for the logistic distribution based on some characterization

Ilia Ragozin[1]

[1] *Department of Mathematics and Mechanics of Saint–Petersburg State University, Universitetsky pr. 28, Stary Peterhof 198504, Russia*

Thursday
August 1st
9:55–10:20

Abstract

There were built two location-free goodness-of-fit tests for the logistic distribution based on a recent characterization by Hua and Lin [1]. The test statistics are based

on suitable functionals of U-empirical distribution functions. One of them has the integral structure, the second one is of Kolmogorov type. For every test statistic there was described the large deviation asymptotics under the null-hypothesis. Then the local Bahadur efficiency was calculated for some alternatives. For instance, in case of scale alternative this efficiency turned out to be 0.837 for the integral test and 0.353 for the Kolmogorov one. Conditions of local optimality in Bahadur sense were also studied.

Keywords: Bahadur efficiency, logistic distribution, large deviations, goodness-of-fit tests, U-statistics.
AMS subject classifications: 62G10, 62G20, 62G30, 62E10.
Acknowledgements: I would like express my gratitude to my scientific advisor Prof. Ya. Yu. Nikitin for interesting problem statement and many useful advice.

Bibliography

[1] Chin-Yuan Hua, Gwo Dong Lin. *Characterizations of the logistic and related distributions*. Journ. of Mathem. Anal. and Appl, 463(2018), N 1, 79-92.

Optimal tests for elliptical symmetry against skew-elliptical alternatives

Thursday
August 1st
10:20–10:45

Slađana Babić[1], Laetitia Gelbgras[2], Marc Hallin[2] and Christophe Ley[3]

[13] *Ghent University*
[2] *Université libre de Bruxelles, ECARES*

Abstract

New efficient tests for elliptical symmetry against skew-ellipticity are presented. The backbone of our construction is the Le Cam theory of asymptotic experiments, and optimality is to be understood in the Le Cam sense, namely asymptotically (in the sample size) and locally (against local generalized skew-elliptical deviations from elliptical symmetry). In each scenario (location specified and unspecified), we first build optimal parametric tests by assuming a given elliptical distribution. Then we render these tests semi-parametric so that they are valid under the entire family of elliptically symmetric distributions we are considering, yet each test will inherit optimal from its parametric antecedent under the parametric distribution the latter was built. We obtain quite astonishing and powerful results. When the location is specified, the optimal parametric tests all are of the same form, meaning that the test statistics do not involve the expression of the density. Consequently, they all lead to the same semi-parametric test, which thus happens to be uniformly

optimal under the entire null hypothesis. This is an extremely rare setting, not only for testing elliptical symmetry, but in general when devising efficient tests. When the location is unspecified, the uniform optimality property logically gets lost, but we still obtain very simple and fast-to-compute test statistics that do not require estimating a density, as is often the case.

Keywords: Semi-parametric inference, elliptical symmetry, skewness, hypothesis testing
AMS subject classifications: 62H15, 62H10

On asymptotic normality of certain linear rank statistics

Viktor Skorniakov[1]

[1] *Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225, Vilnius, Lithuania*

Thursday
July 30th
10:45–11:10

Abstract

We consider asymptotic normality of linear rank statistics under binary randomization rules. Motivated by the set of open problems posed in [2] (see also [3]), our exposition is tightly attached to the context of randomized clinical trials (RCT). In particular, we show that the results obtained extend the ones given in [4] and [6], and apply to a wide class of randomization schemes considered in [1]. However, our derivations rely on some general limit theorem given in [5], and attachment to the RCT is just for the sake of a good demonstration. That is, the obtained results equally well apply to situations where randomization of similar kind takes place: optimal design theory, information theory, etc.

Keywords: Randomization, Asymptotic normality, Linear rank statistics, Clinical trial
AMS subject classifications: 62G10; 62G20; 62P10
Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Yet I would like to thank the managing stuff of my home department in the Faculty of Mathematics and Informatics of Vilnius University as well as the Organizing Committee of the 21st EYSM conference for providing me with an opportunity to participate in the event and present the result.

Bibliography

[1] Baldi Antognini A. and Zagoraiou M. (2015). On the almost sure convergence of adaptive allocation procedures *Bernoulli*, 21 (2), 881–908.
[2] Rosenberger W.F. and Lachin J.M. (2002). *Randomization in Clinical Trials: Theory and Practice*, John Wiley and Sons, New York; Chichester.
[3] Rosenberger W.F. and Lachin J.M. (2016). *Randomization in Clinical Trials: Theory and Practice, Second Edition*, John Wiley and Sons, New Jersey.
[4] Rosenberger W.F. and Rukhin A.L. (2003). Bias properties and nonparametric inference for truncated binomial randomization *J. Nonparametr. Stat.*, 15 (45), 455–465.
[5] McLeish D.L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* 2, 620–628.
[6] Zhang Y. and Rosenberger W.F. (2005). On linear rank tests for truncated binomial randomization *Statist. Probab. Lett.*, 72 (1), 83–92.

New class of suprem-type exponentiality tests based on V-empirical Laplace transforms and Puri-Rubin characterization

Thursday
August 1st
12:40–13:05

Marija Cuparić[1], Bojana Milošević[1] and Marko Obradović[1]

[1]*Faculty of Mathematics, University of Belgrade*

Abstract

We propose a new class of scale free goodness-of-fit tests for the exponential distribution based on Puri-Rubin characterization. For the construction of test statistics we employ weighted L^∞ distance between V-empirical Laplace transforms of random variables that appear in characterization. We derive asymptotic properties, and to assess their quality, we calculate approximate Bahadur efficiency for common close alternatives. For small sample sizes, a simulated power study is performed. The tests are shown to be very efficient and powerful in comparison to many other competitors tests.

Keywords: asymptotic efficiency, Laplace transform, V-statistic with estimated parameters
AMS subject classifications: 62G10, 62G20
Acknowledgements: The work of M. Cuparić and B. Milošević is supported by the MNTRS, Serbia under Grant No. 174012

Distributional Characterizations for Non-normalized Density Functions and Their Applications

Steffen Betsch[1] **and Bruno Ebner**[1]

[1] *Institute of Stochastics, Karlsruhe Institute of Technology*

Thursday
August 1st
13:05–13:30

Abstract

A popular way to construct methods for statistical inference is to craftily employ functional characterizations of probability distributions. In this spirit, the mere fact that a distribution is uniquely determined by its distribution function is used to construct goodness-of-fit tests, like the Kolmogorov-Smirnov test, estimation methods, such as minimum-Crámer-von Mises-distance estimators, or scoring rules to evaluate probabilistic forecasts, like the continuous ranked probability score CRPS. Similar methods are based on characteristic functions, Laplace transforms, quantile functions, and their like. The corresponding procedures, however, fail in situations where only the functional shape of the statistical model is known, that is, if the normalization (or integration) constant of the density function, say, cannot be calculated or handled properly. Models of this type occur in such applied areas as image modeling, signal processing, and machine learning. Although some methods that tackle this problem, like the score matching approach for the estimation of parameters, are available, a broader framework for the construction of further tools seems beneficial.

In this regard, the distributional characterizations we establish in our work [2] provide a reasonable starting point. Extending the notion of the zero-bias distribution often encountered in the context of Stein's method, these characterizations facilitate the construction of suitable statistics, for they do not depend on the normalization constant of the underlying model. The transformations cover any univariate model which admits a continuously differentiable density function that satisfies some weak regularity conditions, and, both in their practical and theoretical handle, the resulting methods resemble classical ones. We already use this approach to construct goodness-of-fit tests (see [1], [3]), and we work on a method for the estimation of parameters. A promising feature of our approach is that even though it is meant to be applied to more complicated statistical models, and despite its wide range of applicability, it appears to perform quite as good as any other procedure even for simpler probability distributions.

Keywords: Distributional characterizations, Goodness-of-fit tests, Non-normalized models, Parameter estimation, Stein's method
AMS subject classifications: 62E10, 60E10, 62F03, 62F10

Bibliography

[1] Betsch, S. and Ebner, B. (2019). A new characterization of the Gamma distribution and associated goodness-of-fit tests. *Metrika*, https://doi.org/10.1007/s00184-019-00708-7.
[2] Betsch, S. and Ebner, B. (2018). Characterizations of continuous univariate probability distributions with applications to goodness-of-fit testing. *ArXiv e-prints*, 1810.06226.
[3] Betsch, S. and Ebner, B. (2019). Testing normality via a distributional fixed point property in the Stein characterization. *TEST*, https://doi.org/10.1007/s11749-019-00630-0.

Factor Models for Functional Time Series in High Dimensions: Representation Theory and Consistent Estimation

Thursday
August 1st
15:00–15:25

Marc Hallin[1], Gilles Nisol[2] and Shahin[3] Tavakoli

[1] *Université libre de Bruxelles*
[2] *Katholiele Universiteit te Leuven*
[3] *University of Warwick*

Abstract

In this paper, we set up theoretical foundations for factor models for a panel of time series of functions. We first show that if the first r eigenvalues of the covariance operator of the cross-section of n FTS are unbounded as n diverges and if the $(r+1)$th eigenvalue is bounded, then we can represent each FTS as a sum of a common component driven by r factors and an idiosyncratic component. We then turn to the estimation of the factors themselves. We show that under conditions weaker than commonly used, we can estimate the space spanned by the factors consistently, thus strengthening existing results for scalar factor models. We also show the consistency of our estimators of the factor loadings and the common component. We then show how these estimators perform on simulated data.

Keywords: Functional Time Series, factor models, high-dimensional statistics.
AMS subject classifications: 62F02.

Order Restricted Inference in Chronobiology

Yolanda Larriba[1], Cristina Rueda[1], Miguel A. Fernández[1] and Shyamal D. Peddada[2]

[1]*Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Spain*
[2]*Department of Biostatistics, Public School of Health, University of Pittsburgh, USA*

Thursday
August 1st
15:25–15:50

Abstract

Blood pressure, body temperature or circadian gene-expressions are just a few of the biological phenomena exhibiting oscillatory processes in nature. Such processes display periodic up-down-up patterns, or rhythms, along periods of time, usually of 24 hours length. The study of these temporal rhythms and how they change under different conditions is called chronobiology [1]. This work is motivated by applications in chronobiology, where researchers are typically interested in discovering the components (e.g. genes) of the oscillatory systems (e.g. circadian clock) that display temporal rhythmic patterns. The contributions of the work are twofold. First, a methodology is developed based on a *circular signal* plus error model that is defined using order restrictions [2]. This mathematical formulation of rhythmicity is simple, easily interpretable and very flexible, with the latter property derived from the non-parametric formulation of the signal. Second, we address various commonly encountered problems in the analysis of oscillatory systems data. Specifically, we propose solutions for the problems of detecting rhythmic signals in an oscillatory system [3], and for the question of estimating the timing of samples when this timing is unknown, such as when tissues are obtained from human biopsies [4]. The proposed methodology is computationally efficient, outperforms the already published ones and is broadly applicable to address a wide range of questions related to oscillatory systems.

Keywords: Constrained Inference, Circular Data, Rhythmicity Detection, Timing Estimation, Oscillatory Systems.
AMS subject classifications: 62P10.
Acknowledgements: The authors gratefully acknowledge the financial support received by the Spanish Ministerio de Ciencia e Innovación and European Regional Development Fund; Ministerio de Economía y Competitividad grant [MTM2015-71217-R to CR and MF] and Spanish Ministerio de Educación, Cultura y Deporte [FPU14/04534 to YL].

Bibliography

[1] Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling.* 11, 11–16.

[2] Silvapulle, M.J. and Sen, P.K. (2005). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*, John Wiley and Sons, New Jersey.
[3] Larriba, Y., Rueda, C., Fernández, M.A. and Peddada, S.D. (2016). Order restricted inference for oscillatory systems for detecting rhythmic signals. *NAR.* 44, e163.
[4] Anafi, R.C., Francey, L.J., Hogenesch, J.B. *et al.* (2017). CYCLOPS reveals human transcriptional rhythms in health and disease. *PNAS.* 114, 5312–5317.

Linear filtering of Gaussian processes in the space of continuous functions

Thursday
August 1st
15:50–16:15

Vít Kubelka[1] **and Bohdan Maslowski**[2]

[1] *Charles University, Prague, Czech Republic*
[2] *Charles University, Prague, Czech Republic*

Abstract This talk deals with an infinite - dimensional generalization of Kalman-Bucy Filter which is a continuous time counterpart to the discrete time linear Kalman Filter. Therefore, it deals with dynamical system described by stochastic differential equations.
First, the continuous time linear filtering problem in finite - dimensional space will be introduced. Afterwards, an extension for Gaussian signal with values in the space of continuous functions and finite - dimensional observation process will be shown and some interesting examples will be discussed, e.g. the signal processes described by linear stochastic partial differential equations driven by Fractional Brownian motion.

Keywords: Kalman - Bucy filter, Stochastic evolution equations, Gaussian processes
AMS subject classifications: 60H05, 60H15, 60G35

Forecasting time series in the light of recent advances in linear mixed modeling and convex optimization

Thursday
August 1st
16:35–17:00

Andrej Gajdoš[1] **and Martina Hančová**[1]

[1]*P. J. Šafárik University, Košice, Slovakia*

Abstract

Prediction of time series within the framework of kriging lays on the idea of the "plug-in" best linear unbiased predictor, the so-called empirical BLUP (EBLUP), where the unknown variance parameters of a chosen time series model are replaced by their estimates generally computed by numerical methods. In our case of kriging ([2]), we consider a class of finite discrete spectrum linear regression models of time series (FDSLRM) introduced by Štulajter ([4]), whose observations are given by a linear mixed model (LMM).

For purposes of real time-series data analysis, forecasting and computational research (Monte Carlo and bootstrap methods), we started to build our own R package (https://github.com/gajdosandrej/fdslrm) on current LMM R packages *nlme, sommer* available at https://cran.r-project.org/), and MATLAB package *mixed* (https://www.mathworks.com/). We also implemented diagnostic tools on the base of recent works ([3]) to improve model diagnostic checking.

Inspired by reference works on convex optimization ([1]), we derived important theoretical relations among the existing estimates of FDSLRM variance parameters. At the same time, in contrast to the nonlinear optimization used in noticed LMM packages, we were able to compute these estimates numerically by the latest convex optimization packages involving interior-point methods which are extremely reliable and time efficient. In addition, we also developed a new, very fast and accurate optimization algorithm for computing mentioned estimates of variances in FDSLRM.

In our conference talk, we will present the results of our research in the context of a real time-series data example from the area of cyber security using dynamic and reproducible documents created by open-source Jupyter technology (https://jupyter.org/).

Keywords: Time series, Linear mixed model, Variance components, EBLUP, Convex programming.

AMS subject classifications: 62M20, 62J05, 62J20, 91B84, 90C25, 65Y20.

Acknowledgements: We would like to thank Jozef Hanč for his technical support including the use of Jupyter technology and his principal help with numerical analysis using convex optimization.

This research is supported by the projects APVV–17–0568, VEGA 1/0311/18 and VVGS–PF–2018–792.

Bibliography

[1] Boyd, S. and Vandenberghe, L. (2009). *Convex Optimization.* Cambridge University Press.

[2] Gajdoš, A., Hančová, M. and Hanč, J. (2017). Kriging methodology and its development in forecasting econometric time series. *Statistika: statistics and economy journal 97*, no.1, 59–73.

[3] Singer, J. M., Rocha, F. M. M., and Nobre, J. S. (2017). Graphical Tools for Detecting Departures from Linear Mixed Model Assumptions and Some Remedial Measures. *International Statistical Review, 85*: 290–324.

[4] Štulajter, F. (2003). The MSE of the BLUP in a finite discrete spectrum LRM. *Tatra Mountains Mathematical Publications 26*, no. 1, 125–131.

Bayesian model selection for a family of discrete valued time series models

Thursday
August 1st
17:00–17:25

Panagiota Tsamtsakiri[1] **and Dimitris Karlis**[1]

[1] *Department of Statistics, Athens University of Economics and Business*

Abstract

Models for univariate count time series can be split into two main categories. The first one is known as parameter driven models where the time autocorrelation comes from an underlying latent process for the mean of the discrete process and the second category is the so called observation driven models where the current observations are related to the past observations in some way. This allows for easier construction and estimation of the model but due to the discreteness and the positivity of the counts special treatment is needed.
In this paper we consider the INARCH model, proposed by [1] and detailed in [3] and [2]. The models have a feedback mechanism for the mean process which is related deterministically with its past values together with past observations. Integer Autoregressive Heteroskedastic (INARCH) models belong to the class of observation driven models. While INARCH type models have gained interest, the problem of selecting the order of the terms in the specification of the models is not developed. We aim at contributing to this direction by proposing a Bayesian model selection approach for INARCH models.
In our study we consider that the conditional distribution of Y_t given the past values is a Poisson distribution and mean linked linearly or log-linearly with past values and past observations. We propose a Bayesian approach based on a transdimensional MCMC approach. At the same time we describe Bayesian estimation for INARCH models which has not been attempted so far. A real data application will be given. Simulation evidence to support the usage of the approach is also provided.

Keywords: discrete valued time series; transdimensional MCMC; model selection;
AMS subject classifications: 62M10
Acknowledgements: This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme $\ll$ Human Resources Development, Education and Lifelong Learning $\gg$ in the context of the project " Strengthening Human Resources Research Potential via Doctorate Research " (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

Bibliography

[1] Ferland, R., A. Latour, and D. Oraichi (2006). *Integer-valued garch process*, Journal of Time Series Analysis 27 (6), 923942.
[2] Fokianos, K. and D. Tjøstheim (2011). *Log-linear poisson autoregression*, Journal of Multivariate Analysis 102 (3), 563578
[3] Fokianos,K., Rahbek, A. Tjøstheim, D. (2009), *Poisson autoregression*, Journal of the American Statistical Association, 104, 1430-1439.

On random environment integer-valued autoregressive models – a survey

Petra Laketa[1]

[1] *University of Niš, Faculty of Sciences and Mathematics, Serbia*

Thursday
August 1st
17:25–17:50

Abstract

A survey of random-environment INAR models is presented. There is a motivation for the introduction of the random environment into INAR models. Different models that are defined are considered, and also their mutual properties. Also, the description of problems that arise in estimation and application are given, as well as the approaches that overcome these problems.

Keywords: random environment; INAR
AMS subject classifications: 62M10
Acknowledgements: Author acknowledges the grant of MNTR 174013 for carrying out this research.

Bibliography

[1] Al-Osh, M.A., Aly, E.E.A.A. (1992) First order autoregressive time series with negative binomial and geometric marginals, Commun. Statist. Theory Meth. 21, 2483–2492.
[2] Al-Osh, M.A., Alzaid, A.A. (1987) First-order integer-valued autoregressive (INAR(1)) process, J. Time Ser. Anal. 8, 261–275.
[3] Aly, E.E.A.A., Bouzar, N. (1994) On Some Integer-Valued Autoregressive Moving Average Models, Journal of Multivariate Analysis 50, 132–151.
[4] Alzaid, A. A., Al-Osh, M.A. (1993) Some autoregressive moving average processes with generalized Poisson marginal distributions, Ann. Inst. Statist. Math. 45, 223–232.
[5] Bakouch, H.S., Ristić, M.M. (2010) Zero Truncated Poisson Integer Valued AR(1) Model, Metrika 72(2), 265-280.

[6] Laketa, P.N., Nastić, A.S., Ristić, M.M. (2018) Generalized Random Environment INAR Models of Higher Order, Mediterranean Journal of Mathematics, 15:1, art.num. 9.

[7] Laketa, P.N., Nastić, A.S. (2019) Conditional least squares estimation of the parameters of higher order Random environment INAR models, Facta UniversitatisSeries Mathematics and Informatics , To Appear

[8] Nastić, A.S., Laketa, P.N., Ristić, M.M. (2016) Random Environment Integer-Valued Autoregressive process, J. Time Ser. Anal. 37, 267–287.

[9] Nastić, A.S., Laketa, P.N., Ristić, M.M. (2019) Random environment INAR models of higher order,RevStat:Statistical Journal, 17:1, 3565.

[10] Ristić, M.M., Bakouch, H.S., Nastić, A.S. (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, J. Stat. Plan. Inference 139, 2218–2226.

[11] Tang M, Wang Y. 2014. Asymptotic behavior of random coefficient INAR model under random environment defined by difference equation. Advances in Difference Equations 2014(1): 19.

Record times of stationary regularly varying time series

Friday
August 2nd
9:30–9:55

Bojan Basrak[1], Hrvoje Planinić[1] and Philippe Soulier[2]

[1] *University of Zagreb, Croatia*
[2] *Université Paris Nanterre, France*

Abstract

In this talk we present a new type of the so–called complete point process convergence result for a stationary, regularly varying and weakly dependent time series $(X_i)_{i\in\mathbb{Z}}$. Due to dependence, extremal observations of (X_i) appear in clusters and the main novelty is that this convergence preserves the information about the temporal ordering of observations belonging to the same cluster. The key tool is the so–called tail process of (X_i). As an application, we deduce the asymptotic behavior of rescaled record times of (X_i), extending the well–known result in the i.i.d. case. These results are based on the joint work with Bojan Basrak and Philippe Soulier [1].

Keywords: Point process, Tail process, Record times
AMS subject classifications: 60G55, 60G70

Bibliography

[1] Basrak, B., Planinić, H., Soulier, P. (2018). An invariance principle for sums and record times of regularly varying stationary sequences. *Probab. Theory Related Fields*, 172(3-4):869–914.

On blind source separation under martingales: A probability theoretic perspective

Niko Lietzén[1]

[1] *Aalto University School of Science, Department of Mathematics and Systems Analysis, Helsinki, Finland*

Friday
August 2nd
9:55–10:20

Abstract

In this work, we consider a blind source separation estimator that is suitable for a subclass of non-stationary processes. In particular, we consider linear blind source separation for martingales. The blind source separation problem is an occurring and a widely studied topic among especially the signal processing and statistics communities. Usually, the goal in blind source separation is to reverse the effects of an unknown mixing system and to recover some signals of interest. There exists a vast number of applications where blind source separation is utilized, e.g., audio applications, biomedical applications and telecommunications, see [1] for a collection.

Under processes that are not i.i.d., the study of asymptotic properties of different blind source separation estimators has received surprisingly little attention, when taking into account the popularity and the number of applications these estimators have. The asymptotic properties of non-i.i.d. blind source separation estimators have been previously considered under very specific classes of stationary processes in, e.g., [2, 3]. To our knowledge, limiting properties of blind source separation estimators have not yet been considered under non-stationary processes.

Our objective is to present some asymptotic properties for a blind source separation estimator under a subclass of non-stationary stochastic processes. We assume that the latent processes of interest have enough structure for us to employ the rich limiting theory of martingales. Martingales are widely applied in financial modeling. We hope to elevate the interest of the blind source separation community towards the financial applications by demonstrating that valid mathematical foundations can be constructed under models that have previously received little attention.

Keywords: blind source separation, convergence, joint diagonalization, martingales

AMS subject classifications: 60G42, 62H12, 94A12

Acknowledgements: Niko Lietzén gratefully acknowledges financial support from the Emil Aaltonen Foundation (grant 180144). The author would also like to thank Pauliina Ilmonen and Lauri Viitasaari for providing their insights, which have helped improve the paper considerably.

Bibliography

[1] Comon, P. and Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, Oxford.

[2] Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2012). Statistical properties of a blind source separation estimator for stationary time series. *Statistics & Probability Letters.* 11, 1865–1873.
[3] Miettinen, J., Nordhausen, K., Oja, H. and Taskinen, S. (2014). Deflation-based separation of uncorrelated stationary time series. *Journal of Multivariate Analysis.* 123, 214–227.

On decomposable multi-type Bellman-Harris branching process for modeling cancer cell populations with mutations

Friday
August 2nd
10:20–10:45

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

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

Abstract

Metastasis, the spread of cancer cells from a primary tumour to secondary location(s) in the human organism, is the ultimate cause of death for the majority of cancer patients. That is why, it is crucial to understand metastases and their evolution in order to successfully combat the disease.
We consider a metastasized cancer cell population after some medical treatment (e.g. chemotherapy). Arriving in a different environment the cancer cells may change their characteristics concerning lifespan and reproduction, thus they may differentiate into different types. Even if the treatment is effective (resulting in subcritical reproduction of all cancer cell types), however, it is possible during cell division for mutations to occur. These mutations can produce a new cancer cell type that is adapted to the treatment (having supercritical reproduction). Cancer cells from this new type may lead to a non-extinction process.
As a continuation of [3] we model the above scenario with a decomposable multi-type Bellman-Harris branching process. Expanding [1] and [2] we investigate relevant quantities such as the probability of extinction of the process until time t and as t approaches infinity, the number of occurred supercritical mutants until time t and as t approaches infinity, 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 also propose numerical schemes for performing calculations.

Keywords: Mutations, Decomposable multi-type branching process, Probability of extinction, Waiting time to escape mutant, Immediate risk of escaping extinction
AMS subject classifications: 2010 Mathematics Subject Classification: Primary 60J80, Secondary 62P10

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.

Bibliography

[1] K. Vitanov, M. Slavtchova-Bojkova, Multitype branching processes in continuous time as models of cancer, *Annuaire de l'Universite de Sofia "St. Kl. Ohridski", Fac. Math and Inf.*, 104, 2017, 193-200.

[2] M. Slavtchova-Bojkova, K. Vitanov, Modelling cancer evolution by multi-type age-dependent branching processes, *Compt. rend. de l'acad. bulgare des Sci.* 71, 2018, 1297–1305.

[3] M. Slavtchova-Bojkova, P. Trayanov, S. Dimitrov, Branching processes in continuous time as models of mutations: Computational approaches and algorithms, *Computational Statistics and Data Analysis*, 113, 2017, 111–124.

Limit Distribution for Some Iterated Partial Summations

Lívia Leššová[1]

[1] *Faculty of Mathematics, Physics and Informatics of the Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava Slovakia*

Friday
August 2nd
10:45–11:10

Abstract

The univariate partial summations were defined in general in [3] as

$$P_x^{(1)} = c_1 \sum_{j=x}^{\infty} g(j) P_j^*, \qquad x = 0, 1, 2, ..., \tag{3}$$

where $\{P_j^*\}_{j=0}^{\infty}$ (parent) and $\{P_x^{(1)}\}_{x=0}^{\infty}$ (descendant) are univariate discrete distributions, $g(j)$ is a real function and c_1 is a normalizing constant. Now we take the descendant of the first generation $\{P_x^{(1)}\}_{x=0}^{\infty}$ from the partial summation (3) as the parent of the second generation while function $g(j)$ remains unaltered, with a proper normalizing constant c_2. We obtain as the result another probability distribution - the descendant of the second generation $\{P_x^{(2)}\}_{x=0}^{\infty}$. Similarly the descendant of the n-th generation is obtained by applying summation (3) to the parent which is the descendant of the $(n-1)$-st generation, with the same function $g(j)$ and with a proper normalizing constant c_n.

The question is whether there exists a limit distribution $\{P_x^{(\infty)}\}_{x=0}^{\infty}$ of such repeated summations for some given parent $\{P_j^*\}_{j=0}^{\infty}$ and some given function $g(j)$. The

existence of this limit distribution in the univariate case for $g(j) = c$ was proved in [4] (the limit distribution is geometric for a wide class of parent distributions). For some types of parental distributions defined on a finite support it is possible to find the limit distribution using the power method (see [1]).
An extension to the bivariate case is possible. (see [2])

Keywords: Dicrete probability distributions, partial summations, limit distribution.
AMS subject classifications: 60E05.
Acknowledgements: Supported by grant VEGA 2/0054/18.

Bibliography

[1] Koščová M., Harman R. and Mačutek J. (2018). Iterated partial summations applied to finite-support discrete distributions. *https://arxiv.org/pdf/1901.08968.pdf* (accessed on 26-Feb-2019).
[2] Leššová L., Mautek J. (2018). On the limit behaviour of finite-support bivariate discrete probability distributions under iterated partial summations. *Communications in Statistics Theory and Methods* (submitted paper).
[3] Mačutek J. (2003). On Two Types of Partial Summations. *Tatra Mountains Mathematical Publications 26,* 403–410.
[4] Mačutek J. (2006). A Limit Property of the Geometric Distribution. *Theory of Probability and its Applications 50(2),* 316–319.

Author index

Affiliation and Contacts

KEYNOTE SPEAKER	AFFILIATION	EMAIL ADDRESS
Ana Colubi	University of Giessen, Germany	ana.colubi@gmail.com
M. Dolores Jiménez Gamero	Universidad de Sevilla, Spain	dolores@us.es
Pavle Mladenović	University of Belgrade, Serbia	paja@matf.bg.ac.rs
Igor Prünster	Bocconi University, Milan, Italy	igor.pruenster@unibocconi.it
Peter J. Rousseeuw	KU Leuven, Belgium	peter@rousseeuw.net

COUNTRY	PARTICIPANT	AFFILIATION	EMAIL ADDRESS
Armenia	Arshak Minasyan	Yerevan State University, Armenia	arsh.minasyan@gmail.com
Austria	Thomas Kuenzer	Graz University of Technology, Austria	kuenzer@TUGraz.at
	Christian Zwatz	University of Vienna, Austria	christian.zwatz@univie.ac.at
Belgium	Gilles Nisol	Katholiele Universiteit te Leuven, Belgium	gilles.nisol@kuleuven.be
	Julien Remy	Universit Libre de Bruxelles, Belgium	julien.remy@ulb.ac.be
Bulgaria	Kaloyan Vitanov	Sofia University ”St. Kliment Ohridski”, Bulgaria	kalovitanov@gmail.com
Croatia	Hrvoje Planinić	University of Zagreb, Croatia	planinic@math.hr
	Una Radojičić	J.J. Strossmayer University of Osijek, Croatia	uradojic@mathos.hr
Cyprus	Andreas Makrides	University of Rouen, France	andreasmaths@hotmail.com
	Nuriye Sancar	Near East University, Nicosia, Cyprus	nuriye.sancar@neu.edu.tr
Czechia	Vít Kubelka	Charles University, Prague, Czechia	kubelka@karlin.mff.cuni.cz
	Daniela Kuruczová	Masaryk University, Czechia	daniela.kuruczova@mail.muni.cz
Denmark	Morten Overgaard	Aarhus University, Denmark	moov@ph.au.dk
	Helene Charlotte Rytgaard	University of Copenhagen, Denmark	hely@sund.ku.dk
Finland	Niko Leitzén	Aalto University School of Science, Helsinki, Finland	niko.lietzen@aalto.fi
	Jenni Niku	University of Jyväskylä, Finland	jenni.m.e.niku@jyu.fi

COUNTRY	PARTICIPANT	AFFILIATION	EMAIL ADDRESS
Germany	Steffen Betsch	Karlsruhe Institute of Techology, Germany	steffen.betsch@student.kit.edu
	Anne van Delft	Ruhr University Bochum, Germany	anne.vandelft@ruhr-uni-bochum.de
Greece	Christina Parpoula	University of the Aegean, Greece	parpoula.ch@gmail.com
	Panagiota Tsamtsakiri	Athens University of Economics and Business, Greece	ptsamtsakthess@gmail.com
Hungary	Ashish Kumar	Eötvös Loránd University, Budapest, Hungary	ashish@cs.elte.hu
	Blanka Szeitl	Bolyai Institute, University of Szeged, Hungary	szeitl@math.u-szeged.hu
Italy	Federico Camerlenghi	University of Milano - Bicocca & Collegio Carlo Alberto, Italy	federico.camerlenghi@unimib.it
	Claudia Di Caterina	Free University of Bozen-Bolzano, Italy	dicaterina@stat.unipd.it
Lithuania	Viktor Skorniakov	Vilnius University, Lithuania	viktor.skorniakov@mif.vu.lt
Netherlands	Alisa Kirichenko	Centrum Wiskunde en Informatica, Amsterdam, the Netherlands	alisa@cwi.nl
Norway	Céline Cunen	University of Oslo, Norway	cmlcunen@math.uio.no
	Vera Đorđilović	University of Oslo, Norway	vera.djordjilovic@medisin.uio.no
Poland	Mariusz Kubkowski	Warsaw University of Technology, Poland	m.kubkowski@ipipan.waw.pl
	Łukasz Rajkowski	University of Warsaw, Poland	l.m.rajkowski@gmail.com
Portugal	Daniela Correia	University of Porto, Portugal	danielamc@med.up.pt
	Luísa Novais	Univeristy of Minho, Portugal	luisa_novais92@hotmail.com
Russia	Ilia Ragozin	Saint–Petersburg State University, Russia	ragza@yandex.ru
	Nikita Zhivotovskiy	Technion I.I.T., Israel	n.zhivotovskiy@skoltech.ru
Serbia	Marija Cuparić	University of Belgrade, Serbia	marijar@matf.bg.ac.rs
	Petra Laketa	University of Niš, Serbia	petralaketa49@gmail.com
Slovakia	Andrej Gajdoš	P. J. Šafárik University, Košice, Slovakia	andrej.gajdos@student.upjs.sk
	Lívia Leššová	Comenius University in Bratislava, Slovakia	livia.lessova@fmph.uniba.sk
Slovenia	Slađana Babić	Ghent University, Belgium	sladana.babic@ugent.be
Spain	Javier Álvarez-Liébana	University of Oviedo, Spain	alvarezljavier@uniovi.es
	Yolanda Larriba	Universidad de Valladolid, Spain	yolanda.larriba@uva.es
Switzerland	Dario Azzimonti	Dalle Molle Institute for Artificial Intelligence, Lugano, Switzerland	dario.azzimonti@idsia.ch
	Alexandre Mösching	University of Bern, Switzerland	alexandre.moesching@stat.unibe.ch
Turkey	Busenur Kızılaslan	Marmara University, Istanbul, Turkey	busenur.sarica@marmara.edu.tr
	Ayça Pamukcu	Acıbadem Mehmet Ali Aydınlar University, İstanbul, Turkey	aycapamukcu9@gmail.com
Ukraine	Stanislav Lohvinenko	Taras Shevchenko National University of Kyiv, Ukraine	stanislav.lohvinenko@gmail.com
United Kingdom	Michele Nguyen	University of Oxford, United Kingdom	michele.nguyen@bdi.ox.ac.uk
	Jure Vogrinc	University of Warwick, United Kingdom	Jure.Vogrinc@warwick.ac.uk

Diamond sponsors

Gold sponsors

Silver sponsors

CIP– Каталогизација у публикацији –
Народна библиотека Србије, Београд

519.2(048)

EUROPEAN Young Statisticians Meeting (21; 2019; Beograd)
Book of Abstracts / 21st European Young Statisticians Meeting, 29 July–2 August, Belgrade, Serbia ; eds. Bojana Milošević, Marko Obradović. – Belgrade : Faculty of Mathematics,2019 (Beograd : Donat graf). – XVI, 63 str. ; 25cm

Na nasl. str.: Bernoulli Society for Mathematical Statistics and Probability. - Tiraž 76. - Bibliografija uz većinu apstrakata. - Registar.

ISBN 978-86-7589-135-2

а) Математичка статистика - Апстракти

COBISS.SR-ID 277806604