22nd European Young Statisticians Meeting – Book of Abstracts

Published by: Department of Psychology & Department of Sociology, School of Social Science, 136, Syggrou Ave 17671 Athens, Greece

For publisher: Panteion University of Social and Political Sciences

Editors: Andreas Makridis, Fotios S. Milienos, Panagiotis Papastamoulis, Christina Parpoula & Athanasios Rakitzis

Place of publication: Athens, Greece

Year of publication: 2021

ISBN: 978-960-7943-22-4

22nd European Young Statisticians Meeting

6 – 10 September 2021, Athens, Greece

Book of Abstracts

Eds. Andreas Makridis, Fotios S. Milienos, Panagiotis Papastamoulis, Christina Parpoula & Athanasios Rakitzis

Bernoulli Society
for Mathematical Statistics
and Probability

Πάντειον Πανεπιστήμιο Κοινωνικών & Πολιτικών Επιστημών
Panteion University of Social and Political Sciences

Preface

This booklet contains basic information about the **22nd European Young Statisticians Meeting** (22nd EYSM) to be held virtually, in the co-organization of *Panteion University of Social and Political Sciences* (Host Institution, Depts. of Psychology & Sociology), *University of the Aegean* (Dept. of Statistics & Actuarial-Financial Mathematics), *Athens University of Economics & Business* (Dept. of Statistics) from Monday, September 6th to Friday, September 10th, 2021.

There are thirty two European countries participating at the 22nd 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
- [-] fifty seven invited young scientists will hold 20-minutes lectures.

Following the tradition of European Young Statisticians Meetings, there are no parallel sessions. The lectures of invited young scientists are divided into fifteen sessions (three sessions per day), twelve of them having four lectures, and three of them three 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, etc.
- Bayesian inference
- Change-point detection
- Characterizations of probability distributions

- Extreme and record value theory
- Functional statistics
- Goodness-of-fit testing
- High-dimensional statistics
- Markov chain Monte Carlo (MCMC) methods
- 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 Proceedings of the 22nd EYSM. The review process for short papers is organized by the IOC, in the way that the IOC representatives personally act as a referee or propose reviewers for papers of participants they invited.

This booklet, beside all important information about the Conference, contains the scientific program, abstracts of all lectures to be given at the 22nd 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 **22nd European Young Statisticians Meeting** could be found at the Conference website https://www.eysm2021.panteion.gr.

22nd European Young Statisticians Meeting

Co-Organized by

Depts. of Psychology & Sociology, Panteion University of Social & Political Sciences (Host Institution)

Dept. of Statistics & Actuarial-Financial Mathematics, University of the Aegean

Dept. of Statistics, Athens University of Economics & Business

Under the auspices of

Bernoulli Society for Mathematical Statistics and Probability

International Organizing Committee

Conference Structure: keynote lectures, invited lectures.

Conference Language: English

Scientific Program

Monday – September 6, 2021

10:30 - 11:00 **Opening Ceremony**

11:00 - 12:40 **Session 1**
Chairman: Alex Karagrigoriou

11:00 – 11:25	Aimilia Savva *Adaptive rates of contraction based on p-exponential priors*
11:25 – 11:50	Aliaksandr Hubin *Reversible Genetically Modified MCMCs*
11:50 – 12:15	Simone Tiberi *BANDITS: a Bayesian hierarchical model for differential splicing accounting for sample-to-sample variability and mapping uncertainty*
12:15 – 12:40	Marina Riabiz *Optimal Thinning of MCMC Output*

12:40 - 13:40 **Mid-day break**

13:40 - 15:20 **Session 2**
Chairman: Vlad Stefan Barbu

13:40 – 14:05	Joona Karjalainen *Parameter estimators of thinned random intersection graphs with non-binomial community size distributions*
14:05 – 14:30	Jakob Peterlin *Tests for functional form of Linear Mixed effects Models*
14:30 – 14:55	Marina Dietrich *Wild Bootstrap for Counting Process-Based Statistics with Application to Fine-Gray Models*
14:55 – 15:20	Alexandre Jacquemain *The penalized bootstrap Lorenz regression procedure*

15:20 – 15:40 **Short break 1**

15:40 - 16:40 **Keynote Lecture**
Chairman: Fotios Milienos

Ingrid Van Keilegom
KU Leuven, Leuven, Belgium
On a Semiparametric Estimation Method for AFT Mixture Cure Models

16:40 – 17:00 **Short break 2**

17:00 - 18:15 **Session 3**
Chairman: Christina Parpoula

17:00 – 17:25 Murat Ozkut
Reliability assessment of systems with two components via phase-type distributions

17:25 – 17:50 Martin Bladt
Matrix Mittag-Leffler distributions and modeling heavy-tailed risks

17:50 – 18:15 Elena Castilla González
Inference for one-shot device test analysis under log-normal lifetimes

Tuesday – September 7, 2021

11:00 - 12:40 **Session 1**
Chairman: Athanasios Rakitzis

11:00 – 11:25	Peter Knaus *A Bayesian Survival Model for Time-Varying Coefficients and unobserved Heterogeneity*
11:25 – 11:50	Negera Wakgari Deresa *On semiparametric modelling, estimation and inference for survival data subject to dependent censoring*
11:50 – 12:15	Mathilde Sautreuil *Study of neural networks to predict survival in oncology*
12:15 – 12:40	Nicola-Ann Stevens *Application of survival techniques to establish environmental and operational controls on road bridge deterioration*

12:40 - 13:40 **Mid-day break**

13:40 - 15:20 **Session 2**
Chairman: Aglaia Kalamatianou

13:40 – 14:05	Petre Babilua *About one nonparametric estimate of the Bernoulli regression function*
14:05 – 14:30	Krzysztof Rudaś *Linear regression for uplift modeling*
14:30 – 14:55	Andrej Srakar *Combinatorial regression model in abstract simplicial complexes*
14:55 – 15:20	Johan Larsson *Look-Ahead Screening Rules for the Lasso*

15:20 – 15:40 **Short break 1**

15:40 - 16:40 **Keynote Lecture**
Chairman: Fotios Milienos

Narayanaswamy Balakrishnan
McMaster University, Hamilton, Ontario, Canada
Efficient Likelihood-Based Inference for the Generalized Pareto Distribution

16:40 – 17:00 **Short break 2**

17:00 - 18:40 **Session 3**
Chairman: Andreas Makrides

17:00 – 17:25 Eleni Manoli
Effectiveness of Indirect Questioning Techniques - Simulation Results

17:25 – 17:50 Royi Jacobovic
Simple sufficient condition for inadmissibility of Moran's single-split test

17:50 – 18:15 Zeljka Salinger
Generalized Gaussian time series model for EEG data

18:15 – 18:40 Yaniv Tenzer
Testing Independence under Biased Sampling

Wednesday – September 8, 2021

11:00 - 12:40 **Session 1**
Chairman: Christina Parpoula

11:00 – 11:25 Jovita Gudan
Comparison of statistics for testing changed segment in a sample

11:25 – 11:50 Emmanouil-Nektarios Kalligeris
Semi-Markov Regime Switching Modeling for Anomaly Detection in Time-Series Incidence Data

11:50 – 12:15 Leo Huberts
Predictive Monitoring Using Machine Learning Algorithms and a Real-Life Example on Schizophrenia

12:15 – 12:40 Laura Boyle
Monitoring overcrowding in a network of hospital emergency departments

12:40 - 13:40 **Mid-day break**

13:40 - 15:20 **Session 2**
Chairman: Panagiotis Papastamoulis

13:40 – 14:05 Saulius Jokubaitis
Sparse structures with LASSO through Principal Components: forecasting GDP components in the short-run

14:05 – 14:30 Mateusz Staniak
Statistical methods for modern proteomics

14:30 – 14:55 Henning Höllwarth
On Estimating the Distribution of the Maximum Likelihood Estimator in Exponential Families

14:55 – 15:20 Chengchun Shi
Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making

15:20 – 15:40 **Short break 1**

15:40 - 16:40 **Keynote Lecture**
Chairman: Athanasios Rakitzis

Christian H. Weiss
Helmut Schmidt University, Hamburg, Germany
On Approaches for Monitoring Categorical Event Series

16:40 – 17:00 **Short break 2**

17:00 - 18:15 **Session 3**
Chairman: Bojana Milošević

17:00 – 17:25 Nikolai Slepov
Distributions transformations and Stein's method

17:25 – 17:50 Mariia Platonova
An analogue of the Feynman-Kac formula for higher order evolution equations

17:50 – 18:15 Ivana Valentić
A CLT for degenerate diffusions with periodic coefficients, and application to homogenization of linear PDEs

Thursday – September 9, 2021

11:00 - 12:40 **Session 1**
Chairman: Fotios Milienos

11:00 – 11:25	Andrea Cappozzo *Penalized model-based clustering for three-way data structures*
11:25 – 11:50	Roi Naveiro *Protecting Classifiers from Attacks*
11:50 – 12:15	Kateřina Koňasová *Supervised classification of replicated point patterns*
12:15 – 12:40	Laura Vana *Analysis of Spatio-Temporal Data Using Bayesian and Formal Methods*

12:40 - 13:40 **Mid-day break**

13:40 - 15:20 **Session 2**
Chairman: Chris Oates

13:40 – 14:05	Janne Kaseva *Quantifying and Assessing Climate Resilience of Plant-Soil Systems Through Response Diversity*
14:05 – 14:30	Jan Vávra *Classification Based on Multivariate Mixed Type Longitudinal Data with an application to the EU-SILC database*
14:30 – 14:55	Steven Golovkine *Clustering multivariate functional data using unsupervised binary trees*
14:55 – 15:20	Yves Staudt *Goodness of Lift in Collision Insurance*

15:20 – 15:40 **Short break 1**

15:40 - 16:40 **Keynote Lecture**
Chairman: Panagiotis Papastamoulis

Sylvia Frühwirth-Schnatter
Vienna University of Economics and Business, Vienna, Austria
From here to infinity – bridging finite and Bayesian nonparametric mixture models in model-based clustering

16:40 – 17:00 **Short break 2**

17:00 - 18:40 **Session 3**
Chairman: Lauri Viitasaari

17:00 – 17:25 Matúš Padyšák
Yield curve modelling in insurance

17:25 – 17:50 Rosanna Cataldo
The Partial Least Squares-Path Modeling for the building of Social Composite Indicators

17:50 – 18:15 Laura Jula Vanegas
Modeling the superposition of dependent binary signals with hidden Markov models

18:15 – 18:40 Máté Szalai
Stochastic modelling of the bactericidal potency

Friday – September 10, 2021

11:00 - 12:40 **Session 1**
Chairman: Milto Hadjikuriakou

11:00 – 11:25 Luigi-Ionut Catana
On the stochastic orders of extremes order statistics

11:25 – 11:50 Zoran Vidović
Posterior properties of the Weibull model for record data

11:50 – 12:15 Nikolaj Thams
Testing in Interventional Distributions

12:15 – 12:40 Beatriz Santos
Star-shaped order for distributions with multidimensional parameters and some applications

12:40 - 13:40 **Mid-day break**

13:40 - 15:20 **Session 2**
Chairman: Athanasios Rakitzis

13:40 – 14:05 Mustafa Hilmi Pekalp
An Application of a Geometric Process Model for Debugging and Testing Costs

14:05 – 14:30 Silius M. Vandeskog
Modelling block maxima with the blended generalised extreme value distribution

14:30 – 14:55 Dragana Radojičić
The Limit Order Book statistical properties

14:55 – 15:20 Jessica Silva Lomba
Mixed moment estimator for space-time heteroscedastic extremes: semi-parametric inference on extreme rainfall

15:20 – 15:40 **Short break 1**

15:40 - 16:40 **Keynote Lecture**
Chairman: Christina Parpoula

Markos V. Koutras
University of Piraeus, Piraeus, Greece
Distributions of statistics related to random samples of random size and pattern occurrences: theory and applications

16:40 – 17:00 **Short break 2**

17:00 - 18:15 **Session 3**
Chairman: Andreas Makrides

17:00 – 17:25 Silvi-Maria Gurova
COVID-19: Study of the spread of the pandemic in Bulgaria

17:25 – 17:50 Petros Barmpounakis
Modelling SARS-CoV2 transmission in a Bayesian framework

17:50 – 18:15 Marian Petrica
A regime switching on Covid-19 analysis and prediction in Romania

18:15 – 19:00 **Closing Ceremony**

Contents

Keynote lectures

On a Semiparametric Estimation Method for AFT Mixture Cure Models

Ingrid Van Keilegom

ORSTAT, KU Leuven, Belgium

Monday
September 6th
15:40–16:40

Abstract

When studying survival data in the presence of right censoring, it often happens that a certain proportion of the individuals under study do not experience the event of interest and are considered as cured. The mixture cure model is one of the common models that take this feature into account. It depends on a model for the conditional probability of being cured (called the incidence) and a model for the conditional survival function of the uncured individuals (called the latency). This work considers a logistic model for the incidence and a semiparametric accelerated failure time model for the latency part. The estimation of this model is obtained via the maximization of the semiparametric likelihood, in which the unknown error density is replaced by a kernel estimator based on the Kaplan-Meier estimator of the error distribution. Asymptotic theory for consistency and asymptotic normality of the parameter estimators is provided. Moreover, the proposed estimation method is compared with several competitors. Finally, the new method is applied to data coming from a cancer clinical trial. This talk is based on the results in [1] and [2].

Keywords: Accelerated failure time model, cure model, Kaplan-Meier estimator, kernel density estimation, semiparametric model.
AMS subject classifications: 62N01, 62N02

Bibliography

[1] Parsa, M. and Van Keilegom, I. (2021a). On a semiparametric estimation method for AFT mixture cure models (in preparation).
[2] Parsa, M. and Van Keilegom, I. (2021b). Accelerated failure time vs Cox proportional hazards mixture cure models: David vs Goliath? (submitted).

Efficient Likelihood-Based Inference for the Generalized Pareto Distribution

N. Balakrishnan

Distinguished University Professor, McMaster University, Canada

Tuesday
September 7th
15:40–16:40

Abstract

The Generalized Pareto Distribution (GPD), introduced originally by Pickands, is widely used for modelling exceedances over thresholds. It is well known that inference for the GPD is a difficult problem since the moments do not all exist for some range of the shape parameter and that the GPD violates the classical regularity conditions in the maximum likelihood method. In this talk, I will describe a novel framework for inference for the GPD, which works successfully for all values of the shape parameter k. I will also present some asymptotic properties of the proposed estimators and related statistics. Based on these results, I will discuss confidence intervals and hypothesis tests. I will then present some Monte Carlo simulation results to demonstrate the overall good performance of the proposed estimators, confidence intervals and hypothesis tests. Finally, I will use two real-data sets to illustrate all the inferential methods discussed.

Keywords: Pareto Distribution, Maximum Likelihood Estimation, Asymptotic Properties of Estimators.
AMS subject classifications: 62F10, 62F12

Bibliography

[1] Nagatsuka, H. and Balakrishnan, N. (2021). Exact likelihood-based inference for the generalized Pareto distribution, *Annals of the Institute of Statistical Mathematics*, to appear.

On Approaches for Monitoring Categorical Event Series

Christian H. Weiß[1]

[1]*Department of Mathematics and Statistics, Helmut Schmidt University, 22043 Hamburg, Germany. E-Mail:* weissc@hsu-hh.de. *ORCID:* 0000-0001-8739-6631.

Wednesday
September 8th
15:40–16:40

Abstract

Methods from statistical process control (SPC) allow to monitor quality-related processes as they occur, for example, in manufacturing and service industries as well as in health surveillance. Here, the most well-known SPC tool is the control chart, where certain quality statistics are computed sequentially in time and used to decide about the actual state of the process [2]. In many manufacturing applications, the monitoring of categorical event series $(X_t)_{t\in\mathbb{N}=\{1,2,\ldots\}}$ is required, i. e., of processes, where the quality characteristics are measured on a qualitative scale. In this context, the control charts are referred to as attributes charts. We distinguish quality features X_t having a finite range consisting of either unordered but distinguishable categories (nominal data), or categories exhibiting a natural order (ordinal data) [3]. We survey three groups of approaches for this task. First, the categorical event series might be transformed into a count process (e. g., event counts, discrete waiting times). After having identified an appropriate model for this count process, diverse control charts are available for the monitoring of the generated counts. Second, control charts might be directly applied to the considered categorical event series, using different charts for nominal than for ordinal data. The latter distinction is also crucial for the respective possibilities of analyzing and modeling these data. Finally, also rule-based procedures from machine learning might be used for the monitoring of categorical event series, where the generated rules are used to predict the occurrence of critical events. We focus on procedures of episode mining, a discipline that was first considered by [1]. Our comprehensive survey of methods and models for categorical event series, which is based on the recent survey [4], is complemented by two real applications: a nominal event sequence regarding paint defects on ceiling fan covers, and an ordinal event sequence regarding so-called "flash" on toothbrush heads.

Keywords: attributes control charts, count time series, episode mining, nominal time series, ordinal time series.
AMS subject classifications: 62M10

Bibliography

[1] Mannila, H., Toivonen, H., Verkamo, A.I. (1997). Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery* 1(3), 259–289.

[2] Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*, 6th edition, John Wiley & Sons, Inc., New York.
[3] Weiß, C.H. (2018). *An Introduction to Discrete-Valued Time Series*, John Wiley & Sons, Inc., Chichester.
[4] Weiß, C.H. (2021). On approaches for monitoring categorical event series. In Tran (ed.): *Control Charts and Machine Learning for Anomaly Detection in Manufacturing*, Springer, forthcoming.

From here to infinity – bridging finite and Bayesian nonparametric mixture models in model-based clustering

Sylvia Früwirth-Schnatter[1]

[1]*Department of Finance, Accounting, and Statistics*

Thursday
September 9th
15:40–16:40

Abstract

This talk reviews the concept of mixture models and their application for model-based clustering from a Bayesian perspective and discusses some recent developments. Two broad classes of mixture models are available. One the one hand, finite mixture models are employed with a finite number K of components in the mixture distribution. One the other hand, Bayesian nonparametric mixtures, in particular Dirichlet process mixtures, are very popular. These models admit infinitely many mixture components and imply a prior distribution on the partition of the data, with a random number of data clusters. This allows to derive the posterior distribution of the number of clusters given the data which contains useful information regarding unobserved heterogeneity in the data.

One reason for the popularity of Dirichlet process mixtures is the common belief that finite mixture models are different in this regard and by selecting the number K of components in the mixture distribution the number of data clusters is automatically forced to be equal to K. However, recent research in finite mixture models has revealed surprising similarities between finite and Bayesian nonparametric mixture models. It has been shown that also for finite mixtures there exists a pronounced difference between the number of components in the mixture distribution and the number of clusters in the data, in particular, if the mixture model is overfitting. The concentration parameter γ in the Dirichlet prior on the mixture weights is instrumental in this respect and, for appropriate choices of γ, finite mixture models also imply a prior distribution on the partition of the data with a random number of data clusters, see e.g. [1] and [3].

In addition, a prior can be put on the number K of components in the mixture distribution. This allows to infer simultaneously K and the number of data clusters from the data within the framework of generalized mixtures of finite mixtures. This new framework, introduced by [2], encompasses many well-known mixture modelling frameworks, including Dirichlet process and sparse finite mixtures. A generic MCMC sampler (called telescoping sampler) is introduced in [2] that allows straightforward MCMC implementation and avoids the tedious design of moves in common trans-dimensional approaches such as RJMCMC.

Keywords: Dirichlet distribution, Dirichlet process mixtures, RJMCMC
Acknowledgements: This talk is based on joint work with Jan Greve, Bettina Grün, and Gertraud Malsiner-Walli. Supported from the Austrian Science Fund (FWF), grant P28740.

Bibliography

[1] Frühwirth-Schnatter, S. and G. Malsiner-Walli (2019). From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification 13*, 33–64.
[2] Frühwirth-Schnatter, S., G. M. Walli, and B. Grün (2020). Generalized mixtures of finite mixtures and telescoping sampling. arXiv preprint 2005.09918v2.
[3] Greve, J., B. Grün, G. Malsiner-Walli, and S. Frühwirth-Schnatter (2020). Spying on the prior of the number of data clusters and the partition distribution in Bayesian cluster analysis. *arXiv 2012.12337*.

Distributions of statistics related to random samples of random size and pattern occurrences: theory and applications

Friday
September 10th
15:40–16:40

Markos V. Koutras

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

Abstract

In the present talk we discuss the exact distribution of statistics based on samples of random variables, with random sample size. We focus in particular to the case when the sample size is a random variable associated to a pattern waiting time distribution in a sequence of binary or multistate trials. The motivation of the models to be studied stems from several areas of applied science such as actuarial science, Financial Risk Management, Quality Control and Reliability, Biostatistics, Educational Psychology, Engineering etc.
Assume that the random variable T denotes the waiting time for the first occurrence of a specific (simple or composite) pattern in a sequence of binary or multistate trials. Let also $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed discrete random variables which are independent of T. Several theoretical results will be given for the distribution of the sum $S = \sum_{t=1}^{T} Y_t$ and the distribution of the rth largest and the of the rth smallest observation of the sample $Y_1, Y_2, \ldots, Y_T$. A number of examples for specific patterns (e.g. runs or scans) will be presented to illustrate the analytical results developed.
Finally, several extensions of the models and related topics of further research interest will be discussed.

Keywords: Pattern waiting times, compound distributions, random order statistics, run and scan statistics, phase-type distributions
AMS subject classifications: 62E15, 60E05

Bibliography

[1] Balakrishnan N. and Koutras M. V. (2002). *Runs and Scans with Applications.* Wiley, NY.
[2] Glaz, J., Pozdnyakov, V., and Wallenstein, S. (2009). *Scan Statistics: Methods and Applications*, Birkhauser, Boston.
[3] He Q-M. (2014). *Fundamentals of Matrix-Analytic Methods*, Springer.
[4] Koutras, M. V. and Eryilmaz, S. (2017) Compound geometric distribution of order k. *Methodology and Computing in Applied Probability*, 19, 377-393.
[5] Koutras, V. M. and Koutras, M. V. (2020). Exact distribution of random order statistics and applications in risk management. *Methodology and Computing in Applied Probability*, 22, 1539-1558.
[6] Tank F. and Eryilmaz S. (2014). The distributions of sum, minima and maxima of generalized geometric random variables, *Statistical Papers*, 56, 1191-1203.
[7] Koutras, V. M., Koutras, M. V. and Dafnis S. D. (2021). A family of induced distributions, Submitted.

Abstracts

Adaptive rates of contraction based on p-exponential priors

Sergios Agapiou and Aimilia Savva

Department of Mathematics and Statistics, University of Cyprus

Monday
September 6th
11:00–11:25

Abstract

We consider Bayesian approaches to non-parametric models. In particular, we use p-exponential priors, which are priors constructed via random series expansions using distributions with tails between Gaussian and exponential. We will study the frequentist asymptotic performance of the posterior distribution in the infinitely informative data limit, in terms of posterior contraction rates. Priors with exponential rather than Gaussian tails, were shown in [1] to be more suitable for modeling spatially inhomogeneous unknown functions, that is functions that are regular in some part and irregular in some other part of their domain.
Our main objective, is to design procedures which give rise to posteriors contracting at rates which are adaptive in the minimax sense. Specifically, we study p-exponential priors with scaling and regularity hyper-parameters, using empirical Bayes and hierarchical Bayes methods of choosing the hyper-parameters.
We built on the work of Agapiou et al. in [1], who developed the general contraction theory for p-exponential priors with fixed hyper-parameters. The difficulty with these priors is that they are not conjugate to the likelihood, even for simple models like non-parametric regression models. In order to establish posterior contraction rates in the empirical and hierarchical Bayes approaches, we rely on the general theory of Rousseau and Szabó in [2], which does not require explicit knowledge for the marginal likelihood.

Keywords: Adaptive Bayesian non-parametric inference, non-Gaussian priors, hierarchical and empirical Bayes procedures
AMS subject classifications: Primary 62G20; secondary 62G05, 60G50

Bibliography

[1] S. Agapiou, M. Dashti, and T. Helin (2018). Rates of contraction of posterior distributions based on p-exponential priors. *arXiv:1811.12244 (to appear in Bernoulli)*.
[2] J. Rousseau and B. Szabó (2017). Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *The Annals of Statistics* vol.45 no.2, 833–865.

Reversible Genetically Modified MCMCs

Monday
September 6th
11:25–11:50

Aliaksandr Hubin[1], **Geir Storvik**[1] **and Florian Frommlet**[2]

[1]*University of Oslo*
[2]*Medical University of Vienna*

Abstract

In this work, we introduce a reversible version of a genetically modified Markov chain Monte Carlo algorithm (GMJMCMC) for inference on posterior model probabilities in complex functional spaces, where the number of explanatory variables or functions of explanatory variables is prohibitively large for simple Markov Chain Monte Carlo methods. A genetically modified Markov chain Monte Carlo algorithm (GMJMCMC) was introduced in [5, 4, 2] for Bayesian model selection/averaging problems when the total number of function of covariates is prohibitively large. More specifically, these applications include GWAS studies with Bayesian generalized linear models [2] as well as Bayesian logic regressions [5] and Bayesian generalized nonlinear models [4]. If its regularity conditions are met, GMJMCMC algorithm can asymptotically explore all models in the defined model spaces. At the same time, GMJMCMC is not a proper MCMC in a sense that its limiting distribution does not correspond to marginal posterior model probabilities and thus only renormalized estimates of these probabilities [3, 1] can be obtained. Unlike the standard GMJMCMC algorithm, the introduced algorithm is a proper MCMC and its limiting distribution corresponds to posterior marginal model probabilities in the explored model spaces under reasonable regularity conditions.

Keywords: Markov chain Monte Carlo; Mode lumping in MCMC; Genetic algorithms; Bayesian Model selection; Bayesian Model averaging

AMS subject classifications: 62-02, 62-09, 62F07, 62F15, 62J12, 62J05, 62J99, 62M05, 05A16, 60J22, 92D20, 90C27, 90C59

Acknowledgements: We would also like to acknowledge NORBIS (https://norbis.w.uib.no/) for funding academic stay of the first author in Vienna.

Bibliography

[1] M. A. Clyde, J. Ghosh, and M. L. Littman. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, 20(1):80–101, 2011.

[2] A. Hubin. *Bayesian model configuration, selection and averaging in complex regression contexts*. PhD thesis, University of Oslo, 2018.

[3] A. Hubin and G. Storvik. Mode jumping MCMC for Bayesian variable selection in GLMM. *Computational Statistics & Data Analysis*, 127:281–297, 2018.

[4] A. Hubin, G. Storvik, and F. Frommlet. Flexible Bayesian Nonlinear Model Configuration. *arXiv preprint arXiv:2003.02929*, 2020.

[5] A. Hubin, G. Storvik, and F. Frommlet. A novel algorithmic approach to Bayesian logic regression (with Discussion). *Bayesian Analysis*, 15(1): 263-333 (March 2020).

BANDITS: a Bayesian hierarchical model for differential splicing accounting for sample-to-sample variability and mapping uncertainty

Simone Tiberi and Mark D Robinson

Department of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich

Monday
September 6th
11:50–12:15

Abstract

Alternative splicing plays a fundamental role in the biodiversity of proteins as it allows a single gene to generate several transcripts and, hence, to code for multiple proteins. However, variations in splicing patterns can be involved in diseases. When investigating differential splicing (DS) between conditions, typically healthy *vs.* disease or treated *vs.* untreated, scientists are increasingly focusing on differential transcript usage (DTU), i.e. in changes in the proportion of transcripts.
A big challenge in DTU analyses is that, unlike gene level studies, the counts at the transcript level, which are of primary interest, are not observed because most reads map to multiple transcripts. Tools such as salmon or kallisto allow, via expectation-maximization algorithms, to estimate the expected number of fragments originating from each transcript. Most DTU methods follow a plug-in approach and take the estimated counts as input by treating them as real transcript counts, thus neglecting the uncertainty in the estimates. In order to avoid this issue, other DTU methods consider what transcripts each read is compatible with (also called equivalence class); nevertheless, they assume that all samples share the same transcript proportions.
To overcome the limitations of current methods for DTU, we present BANDITS, a method to perform DTU based on RNA-seq data. BANDITS uses a Bayesian hierarchical model, with a Dirichlet-multinomial structure, to explicitly model the variability between samples. Our tool also models mapping uncertainty: it inputs the equivalence class of each read and treats the allocations of reads to the transcripts as latent variables. The parameters of the model are inferred via Markov chain Monte Carlo (MCMC) techniques, and the stationarity of the posterior chains is assessed via Heidelberg and Welch's convergence diagnostic. Despite the computational complexity of full MCMC algorithms, the core of our method is coded in

C++, which makes BANDITS highly efficient and feasible to run on a laptop, even for complex organisms.
Our method tests for DTU at both transcript and gene level, allowing scientists to investigate what specific transcripts are differentially used in selected genes. Furthermore, our tool is not limited to two group comparisons and also allows to test for DTU when samples belong to more than two groups.
We will show how, both in simulation studies and experimental data analyses, the proposed methodology outperforms existing tools.
BANDITS is available as a Bioconductor R package: https://bioconductor.org/packages/BANDITS.

Keywords: Alternative splicing, Differential splicing, RNA-seq, Bayesian hierarchical modelling, MCMC.
AMS subject classifications: 62P10.

Bibliography

[1] Tiberi, S. and Robinson, M. D. (2020). BANDITS: Bayesian differential splicing accounting for sample-to-sample variability and mapping uncertainty. *Genome biology* 21, 1–13.

Optimal Thinning of MCMC Output

Monday
September 6th
12:15–12:40

Marina Riabiz[1,2] , Wilson Ye Chen[3] , Jon Cockayne[2] , Steven A. Niederer[1], Lester Mackey[5] and Chris. J. Oates[1,6]

[1]*King's College London, UK*
[2]*Alan Turing Institute, UK*
[3]*University of Sydney, Australia*
[4]*Oxford University, UK*
[5]*Microsoft Research, US*
[6]*Newcastle University, UK*

Abstract

The use of heuristics to assess the convergence and compress the output of Markov chain Monte Carlo can be sub-optimal in terms of the empirical approximations that are produced. Typically a number of the initial states are attributed to "burn in" and removed [2], whilst the remainder of the chain is "thinned" if compression is also required. In this work we consider the problem of retrospectively selecting a subset of states, of fixed cardinality, from the sample path such that the approximation provided by their empirical distribution is close to optimal. A novel method is proposed, based on greedy minimisation of a kernel Stein discrepancy [4, 3, 1], that is

suitable when the gradient of the log-target can be evaluated and an approximation using a small number of states is required. Theoretical results guarantee consistency of the method and its effectiveness is demonstrated in the challenging context of parameter inference for ordinary differential equations. Software is available at `http://stein-thinning.org/`.

Keywords: Bayesian computation, greedy optimisation, Markov chain Monte Carlo, reproducing kernel, Stein's method.
AMS subject classifications: 62F15.
Acknowledgements: This work was supported by: the Lloyd's Register Foundation programme on data-centric engineering at the Alan Turing Institute, UK; the British Heart Foundation - Alan Turing Institute cardiovascular data science award (BHF; SP/18/6/33805); the Wellcome/EPSRC Centre for Medical Engineering (WT203148/Z/16/Z).

Bibliography

[1] Chwialkowski, K., Strathmann, H. and Gretton, A. (2016). A kernel test of goodness of fit. *Proceedings of the 33rd International Conference on Machine Learning*
[2] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* 7, 457–472
[3] Gorham, J. and Mackey, L. (2017). Measuring sample quality with kernels. *Proceedings of the 34th International Conference on Machine Learning*
[4] Liu, Q. and Lee, J. D. (2017). Black-box importance sampling. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*

Parameter estimators of thinned random intersection graphs with non-binomial community size distributions

Joona Karjalainen

Aalto University School of Science, Department of Mathematics and Systems Analysis, Espoo, Finland

Monday
September 6th
13:40–14:05

Abstract

Many types of data can be naturally represented and modeled as networks. A rich class of statistical network models is obtained when we assume that the links are generated by one or more independent and partly overlapping communities. Specifically, we consider the case where the communities are Bernoulli random graphs with

a common size distribution. When each community forms a complete graph, the model reduces to the widely studied passive random intersection graph [2], which correctly predicts some aspects of real networks, such as the correlation between degrees and clustering [1]. Introducing a nontrivial link probability $q < 1$ within the communities allows us to control certain properties of the model, most obviously the numbers of generated cliques.
Moment-based parameter estimators were introduced in [3] for the case where the sizes of the communities are binomially distributed and the number of communities grows linearly with the number of nodes. In this work we show that a similar approach can also be used for other families of distributions, and demonstrate this with examples. We also discuss sufficient conditions for the consistency of the estimators as the number of nodes tends to infinity.

Keywords: complex networks, random graphs, parameter estimation, network models
AMS subject classifications: 05C80, 05C07, 62F12, 91D30
Acknowledgements: This work was supported by the Magnus Ehrnrooth foundation.

Bibliography

[1] Bloznelis, M. (2013). Degree and clustering coefficient in sparse random intersection graphs. *Ann. Appl. Probab.* 23(3), 1254–1289.
[2] Godehardt, E. and Jaworski, J. (2003). Two models of random intersection graphs for classification. *Stud. Classification, Data Anal. Knowledge Organ.*, 67–81. Springer, Berlin.
[3] Karjalainen, J., van Leeuwaarden, J.S.H., and Leskelä, L. (2018). Parameter estimators of sparse random intersection graphs with thinned communities. In *Proc. 15th Workshop on Algorithms and Models for the Web Graph (WAW)*, 44–64. Springer, Cham.

Tests for functional form of Linear Mixed effects Models

Monday
September 6th
14:05–14:30

Jakob Peterlin, Nataša Kejžar and Rok Blagus

Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana

Abstract

Linear mixed effects models (LMMs) are a popular and powerful tool for analyzing clustered or repeated observations for numeric outcomes. LMMs, as described for

example in [1], consist of a fixed and a random component which are specified in the model through their respective design matrices. Checking if the two design matrices are correctly specified is crucial since misspecifying them can affect the validity and efficiency of the analysis.
Together with Rok Blagus and Nataša Kejžar we figured out how to use certain random processes to test the appropriateness of the assumed design matrices. We show howed these processes can be used to test for the goodness-of-fit of the entire model, or its fixed and/or random component. We derived a procedure for obtaining the p-values which can be based on sign-flipping, simulations or bootstrap and showed, theoretically with the help of [2] as well as by using a large Monte Carlo simulation study, its validity. The proposed test(s) can be used with models with (arbitrary number of) multi-level or crossed random effects and could also be extended to other similar types of models.
I will briefly present the theoretic foundations and their applications to our method. I will also introduce the proposed method, its assumptions and applications.

Keywords: Asymptotic convergence, Correlated data, Empirical stochastic processes, Monte-Carlo simulations, Wild bootstrap
AMS subject classifications: 62J05
Acknowledgements: Slovenian Research Agency, Grant/Award Number: N1-0035, P3-0154

Bibliography

[1] Demidenko, E. (2013), *Mixed models; Theory and Applications with R*, John Wiley and Sons, New Jersey.
[2] van der Vaart A. W. and Wellner J.Ã. (1996), *Weak Convergence and Empirical Processes*, Springer, New York

Wild Bootstrap for Counting Process-Based Statistics with Application to Fine-Gray Models

Marina Dietrich, Dennis Dobler and Mathisca de Gunst

Department of Mathematics, Vrije Universiteit Amsterdam

Monday
September 6th
14:30–14:55

Abstract

We consider the wild bootstrap for counting process-based statistics and choose the estimators involved in the Fine-Gray model as a representative application. In particular, the Fine-Gray model is used in competing risks settings, in which we focus on censoring-complete data. Here, the goal is to construct asymptotically valid

time-simultaneous confidence bands for the cumulative incidence function using the wild bootstrap. Especially for small samples, the flexibility of the wild bootstrap with possibly non-normal multipliers seems preferable over a Gaussian approximation because the latter is only capturing the asymptotic distribution.
We use elegant martingale arguments in order to theoretically justify the asymptotic validity of the wild bootstrap as an approximation procedure. Here, we first consider general counting process-based statistics, and then refine this approach for estimators in the Fine-Gray model. Moreover, we empirically study the sample performance of the wild bootstrap confidence bands for the cumulative incidence function in Fine-Gray models with respect to different types of multipliers and transformations by means of simulations. In addition, we illustrate the developed method by investigating the impact of Pneumonia for intensive care unit patients on the probabilities of alive discharge versus hospital death.

Keywords: Counting processes, Confidence bands, Fine-Gray model, Martingale theory, Wild bootstrap.
AMS subject classifications: 62N02.

Bibliography

[1] Andersen, P. K. and Borgan, Ø. and Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag New York.
[2] Fine, J. P. and Gray, R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association,* Volume 94, Number 446, 496–509.
[3] Lin, D. Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine,* Volume 16, 901–910.

The penalized bootstrap Lorenz regression procedure

Monday
September 6th
14:55–15:20

Cédric Heuchenne[1], Alexandre Jacquemain[2] and Eugen Pircalabelu[3]

[1]*HEC Liège, University of Liège*
[2]*Université catholique de Louvain, ISBA*
[3]*Université catholique de Louvain, ISBA*

Abstract

The explained Gini coefficient, introduced by [1], is a measure of the economic inequality that can be attributed to a set of covariates. Similarly to the R^2, this quantity never decreases as we keep introducing new covariates and may suffer from

overfitting on large datasets. In this paper, we propose a penalized bootstrap procedure which selects the relevant covariates and produces inference on the explained Gini coefficient, while avoiding overfitting. The obtained estimator achieves the Oracle property and can be computed efficiently. In this respect, we introduce the SCAD-FABS algorithm, an adaptation of the FABS algorithm proposed by [2] to the SCAD penalty. The performance of the procedure is assessed via Monte-Carlo simulations. Finally, a real-data example is presented.

Keywords: Lorenz curve, Income inequality, Single-index, SCAD, FABS algorithm.

Bibliography

[1] Heuchenne, C. and Jacquemain, A. (2021). Inference for monotone single-index conditional means: a Lorenz regression approach. *Computational Statistics and Data Analysis*. Under Review.
[2] Shi, X. Huang, Y. Huang, J. and Ma, S. (2018). A Forward and Backward Stagewise algorithm for nonconvex loss functions with adaptive Lasso. *Computational Statistics and Data Analysis* 124, 235–251.

Reliability assessment of systems with two components via phase-type distributions

Murat Ozkut

Izmir University of Economics

Monday
September 6th
17:00–17:25

Abstract

Phase-type distributions have been found to be very useful in reliability and operational research. Their matrix-based representations are mathematically tractable in terms of calculations. This work is concerned with reliability properties of systems consisting of two dependent components. In particular, the lifetimes of the components are assumed to have bivariate discrete phase-type distribution. The reliability properties of series and parallel systems are investigated under this bivariate setting.

Keywords: Bivariate distribution, Phase-type distribution, reliability function, mean residual life function.
AMS subject classifications: 62N05.

Bibliography

[1] Eisele, K.T. (2008). Recursions for multivariate compound phase variables. *Insurance: Mathematics and Economics,* 42, 65–72.
[2] Krishna, H. and Pundir, S. P. (2009). A Bivariate Geometric Distribution with Applications to Reliability. *Communications in Statistics, Theory and Methods,* 38, 1079-1093.
[3] Roy, D. (1993). Reliability Measures in the Discrete Bivariate Set-Up and Related Characterization Results for a Bivariate Geometric Distribution. *Journal of Multivariate Analysis,* 46, 362–373.
[4] Tank, F. and Eryilmaz, S. (2020). On bivariate compound sums. *Journal of Computational and Applied Mathematics,* 365, 112371.

Matrix Mittag–Leffler distributions and modeling heavy-tailed risks.

Monday
September 6th
17:25–17:50

Martin Bladt

University of Lausanne, Switzerland

Abstract

We define the class of matrix Mittag-Leffler distributions and study some of its properties. We show that it can be interpreted as a particular case of an inhomogeneous phase-type distribution with random scaling factor, and alternatively also as the absorption time of a semi-Markov process with Mittag-Leffler distributed inter-arrival times. We then identify this class and its power transforms as a remarkably parsimonious and versatile family for the modeling of heavy-tailed risks, which overcomes some disadvantages of other approaches like the problem of threshold selection in extreme value theory.

Keywords: absorption times, renewal processes, phase–type distributions
AMS subject classifications: 62G32

Bibliography

[1] Albrecher, Hansjörg and Bladt, Martin and Bladt, Mogens (2020). Matrix Mittag–Leffler distributions and modeling heavy-tailed risks. *Extremes.* 23, 425–450.
[2] Albrecher, Hansjörg and Bladt, Martin and Bladt, Mogens (2021). Multivariate Matrix Mittag-Leffler distributions. *Annals of the Institute of Statistical Mathematics.* 73, 369–394.
[3] Albrecher, Hansjörg and Bladt, Martin and Bladt, Mogens (2020). Multivariate fractional phase–type distributions. *Fractional Calculus and Applied Analysis.* 5, 1431–1451.

Inference for one-shot device test analysis under log-normal lifetimes

N. Balakrishnan[1] **and Elena Castilla**[2]

[1] *McMaster University, Hamilton, Canada*
[2] *Complutense University of Madrid, Spain*

Monday
September 6th
17:50–18:15

Abstract

One-shot device testing data, also known as current status data in survival analysis, come from testing one-shot devices that are used only once and get destroyed after use. The only collected information is whether they failed either before or after the inspection time. In the recent years, the study of this kind of devices has been developed under different assumptions of lifetime distributions, such as exponential, gamma and Weibull ([1]). In particular, there is a growing body of literature that focuses their efforts on developing divergence-based robust inference for one-shot device analysis under the cited distributions (see, for example, [2]). However, one-shot device testing analysis under lognormal lifetime distribution has not been studied yet. In this talk, after providing an overview on the existing methods, we develop inference based on one-shot device test data with lognormal distribution. A simulation study is developed to assess the performance of the proposed method and some real-life data are analyzed for illustrative purpose.

Keywords: Lognormal distribution, One-shot devices, Reliability.
AMS subject classifications: 62N02; 62N05.
Acknowledgements: This research is partially supported by Grant PGC2018-095194-B-I00 from *Ministerio de Ciencia, Innovacion y Universidades* (Spain). E. Castilla is member of the *Instituto de Matematica Interdisciplinar*, Complutense University of Madrid.

Bibliography

[1] Balakrishnan, N., Ling, M. H. and So, H. Y. (2021). *Accelerated Life Testing of One-shot Devices: Data Collection and Analysis 1st Edition.* Wiley-Blackwell.
[2] Balakrishnan, N., Castilla, E., Martin N. and Pardo, L. (2020). Robust inference for one-shot device testing data under Weibull lifetime model. *IEEE transactions on Reliability.* **69**(3), 937–953.

A Bayesian Survival Model for Time-Varying Coefficients and unobserved Heterogeneity

Tuesday
September 7th
11:00–11:25

Peter Knaus and Daniel Winkler

Vienna University of Economics and Business

Abstract

Two sources of heterogeneity are often overlooked in the applied survival literature. On the one hand, time-varying hazard contributions of explanatory variables cannot be captured in the widely used Cox proportional hazard model. To this end, this paper investigates a dynamic survival model in the spirit of [3] within a Bayesian framework. Such a specification allows parameters to gradually evolve over time, thus accounting for time-varying effects.
On the other hand, unobserved heterogeneity across (a potentially large number of) groups is often ignored, leading to invalid estimators. This paper makes accounting for such effects feasible for even large numbers of groups through a shared factor model, which picks up unexplained covariance in the error term.
Building on the Markov Chain Monte Carlo scheme of [4] allows the usage of shrinkage priors to avoid overfitting in such a highly parameterized model. This paper uses the triple gamma prior introduced by [2] in the same fashion as [1] to detect which parameters should be included in the model and which should be allowed to vary over time. Finally, an R package which makes the routine easily available is introduced.

Keywords: Bayesian Survival Model, Time-varying Effects, Variable Selection, Factor Model, Software Package
AMS subject classifications: 62N02, 62F15

Bibliography

[1] Cadonna, A., Frühwirth-Schnatter, S., and Knaus, P. (2020). Triple the Gamma-A Unifying Shrinkage Prior for Variance and Variable Selection in Sparse State Space and TVP Models. *Econometrics*, 8(2), 20.
[2] Griffin, J., and Brown, P. (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis*, 12(1), 135-159.
[3] Hemming, K., and Shaw, J. E. H. (2005): A class of parametric dynamic survival models. *Lifetime data analysis*, 11(1), 81-98.
[4] Wagner, H. (2011). Bayesian estimation and stochastic model specification search for dynamic survival models. *Statistics and Computing*, 21(2), 231-246.

On semiparametric modelling, estimation and inference for survival data subject to dependent censoring

Negera Wakgari Deresa and Ingrid Van Keilegom

ORSTAT, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

Tuesday
September 7th
11:25–11:50

Abstract

When modelling survival data, it is common to assume that the survival time T is conditionally independent of the censoring time C given a set of covariates. However, there are numerous situations in which this assumption is not realistic. The goal of this paper is therefore to develop a semiparametric normal transformation model, which assumes that after a proper nonparametric monotone transformation, the vector (T, C) follows a linear model, and the vector of errors in this bivariate linear model follows a standard bivariate normal distribution with possibly non-diagonal covariance matrix. We show that this semiparametric model is identifiable, and propose estimators of the nonparametric transformation, the regression coefficients and the correlation between the error terms. It is shown that the estimators of the model parameters and the transformation are consistent and asymptotically normal. We also assess the finite sample performance of the proposed method by comparing it with an estimation method under a fully parametric model. Finally, our method is illustrated using data from the AIDS Clinical Trial Group 175 study.

Keywords: Association, dependent censoring, nonparametric transformation, survival analysis.
AMS subject classifications: 62N99.
Acknowledgements: The financial support from the European Research Council (ERC) is gratefully acknowledged.

Bibliography

[1] Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* 89, 659–668.
[2] Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.
[3] Deresa, N. W. and Van Keilegom, I. (2020a). Flexible parametric model for survival data subject to dependent censoring. *Biom. J.* 62, 136–156.
[4] Deresa, N. W. and Van Keilegom, I. (2020b). A multivariate normal regression model for survival data subject to different types of dependent censoring. *Comput. Statist. Data Anal.* 144, 106879.
[5] Emura, T. and Chen, Y. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*, Springer.

Study of neural networks to predict survival in oncology

Tuesday
September 7th
11:50–12:15

Mathilde Sautreuil, Sarah Lemler and Paul-Henry Cournède

Laboratory of Mathematics and Informatics (MICS), CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

Abstract

In this talk, we are interested in studying neural networks' potential to predict survival in oncology. In a clinical study in oncology, the number of variables to characterize patients can be huge with, for example, clinical, genomic, and radiology image data. In contrast, the number of patients in cohorts remains relatively small. We are in what is called a high-dimensional framework, that is when the number of variables is much larger than the sample size. A classical model used to deal with survival data is the Cox model [2]. Some regularization procedures have been proposed to deal with this model in high-dimension. However, they prove limited when the dimension becomes too large. Moreover, the proportional Cox model's hypothesis is not always satisfied. Neural networks have provided state-of-the-art models in a lot of research domains. We have contributed to explore their potential in survival analysis, especially in high-dimension. Two strategies have been considered. The first one is based on the Cox model: the neural network is used to replace the linear dependency in the covariates to determine the Cox hazard function. A second and less studied approach directly predicts the discretized hazard function. In our work, we have adapted this method to the high-dimensional setting. In this talk, we present a comparative study of the performances of different models. We compare a neural network based on the Cox model called Cox-nnet [1] with those based on our discrete-time models adapted to the high-dimension, called NNsurv and NNsurv deep. We consider the Lasso procedure [3] applied to the Cox partial log-likelihood as the benchmark. We created a simulation plan to make this comparison more relevant. The data are simulated from different survival models (Cox, Accelerated Failure Time, and Accelerated Hazards models) with varying complexity levels (data satisfying the proportional risks assumptions, data with survival curves crossing...). We also make the sample size and the number of covariates vary, and study the effect of censorship and sparsity. We consider the Concordance index and the Integrated Brier Score to compare performances. Finally, we apply all the methods to two real datasets in oncology. We conclude that the method with the best performances is generally Cox-nnet in the simplest situations. However, in the most complex ones, the deep version of the neural networks directly predicting the discrete risks (NNsurv deep) proves superior. It is notably the case with non-proportional risks and crossing survival curves.

Keywords: Neural networks, survival analysis, high-dimension, Cox model
AMS subject classifications: 62N02

Bibliography

[1] T. Ching, X. Zhu, and L. X. Garmire. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLOS Computational Biology*, 14(4):e1006076, Apr. 2018. ISSN 1553-7358.

[2] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972. ISSN 0035-9246.

[3] R. Tibshirani. The Lasso Method for Variable Selection in the Cox Model. *Statistics in Medicine*, 16(4):385–395, 1997. ISSN 1097-0258.

Application of survival techniques to establish environmental and operational controls on road bridge deterioration

Nicola-Ann Stevens[1], Myra Lydon[1], Adele H Marshall[2] and Su E Taylor[1]

[1] *School of Natural and Built Environment, Queen's University Belfast*
[2] *School of Mathematics and Physics, Queen's University Belfast*

Tuesday
September 7th
12:15–12:40

Abstract

One of the key components of a bridge management system (BMS) is the deterioration model, whose accuracy will determine the quality of future maintenance, rehabilitation and replacement (MR&R) decisions. The current state of the art for deterioration models in commercial BMS is the application of Markov chains to determine the probability of transitioning between condition states. However, in recent years research has moved to looking at alternative approaches due to the assumption of a constant bridge population and stationary transition probabilities. This can compromise the efficacy of the deterioration model for predicting deterioration of complex aging bridge structures.

This research focuses on the application of survival analysis where the survival time is the time spent in each condition state and the 'failure' is transition of the bridge condition to a worse state. Bridge condition states are measured on an ordinal or numerical scale, common examples range from 1-4 or 0-100 for the UK standardised bridge condition index (BCI). This paper presents the application of survival analysis to establish environmental controls and explores the comprehensive assessment of the utilisation of the Cox Proportional Hazards (PH) model. Expanding on the authors previous research which applied survival techniques to identify bridge performance indicators, this research uses live bridge condition data from Northern Ireland, based on over 6000 bridges which form the strategic and regional road network.

Keywords: survival analysis, bridge management systems, Markov chains, deterioration modelling
AMS subject classifications: 62N03
Acknowledgements: The authors would like to thank the Department for Infrastructure (DfI) for the access to the complete bridge management records, the technical support, and allowing the analysis and findings to be used in this paper. The financial support of the Royal Academy of Engineering under the research fellowship program and The Queen's University Belfast Leverage Studentship Scheme are also gratefully acknowledged.

About one nonparametric estimate of the Bernoulli regression function

Tuesday
September 7th
13:40–14:05

Petre Babilua

Department of Mathematics, Faculty of Exact and Natural Sciences, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

Abstract

Let a random variable Y take two values 1 and 0 with probabilities p ("success") and $1-p$ ("failure"). Assume that the probability of "success" p is the function of an independent variable $x \in [0,1]$, i.e. $p = p(x) = P(Y=1|x)$. Assume that $x_k = \frac{k}{n}$, $k = 0, 1, ..., n$, are the points of division of the interval [0,1] and we have Y_i, $i = 0, 1, ..., n$, which are independent Bernoulli random variables with

$$P(Y_i = 1|x_i) = p(x_i), \quad P(Y_i = 0|x_i) = 1 - p(x_i).$$

The problem consists in estimating a function $p(x)$, $x \in [0,1]$, based on sample $Y_0, Y_1, ..., Y_n$. A problem like this one arises, for instance, in biology, also when studying corrosion processes, and so on. As an estimate for $p(x)$ let us consider the following statistic

$$\widehat{p}_n(x) = \sum_{k=0}^{n} Y_k b_k(n,x),$$

where $b_k(n,x) = C_n^k x^k (1-x)^{n-k}$, $k = 0, 1, ..., n$ is a binomial distribution with probability of "success" $p(x)$, $x \in (0,1)$.
Theorem. Let $p(x)$, $x \in [0,1]$ have a bounded derivative of second order. Then
(*a*) $\widehat{p}_n(x)$ is a consistent estimate of $p(x)$ at all points $x \in (0,1)$;
(*b*) $\sqrt{n}(\widehat{p}_n(x) - p(x))\sigma^{-1}(x) \xrightarrow{d} N(0,1)$, $x \in (0,1)$,

where

$$\sigma^2(x) = p(x)(1-p(x))[4\pi x(1-x)]^{-\frac{1}{2}}$$

(where $\xrightarrow{d}$ denotes convergence in distribution and $N(0,1)$ is a random variable that has a standard normal distribution $\Phi(x)$).

Keywords: Bernstein polynomial, Bernoulli regression function, consistency, power of tests.
AMS subject classifications: 62G10, 62G20.

Linear regression for uplift modeling

Krzysztof Rudaś[1,2] **and Szymon Jaroszewicz**[1]

[1] *Institute of Computer Science Polish Academy of Sciences*
[2] *Faculty of Mathematics and Information Science, Warsaw University of Technology*

Tuesday
September 7th
14:05–14:30

Abstract

Uplift modeling is an approach, which allows for predicting the effect of an action (e.g. a marketing campaign or a medical treatment) on a given individual. To achieve this we divide our sample into two subgroups: treatment, which is subjected to the action, and control on which no action is taken. The partitioning of data into treatment and control group is done through randomization. The task then is to estimate the difference between responses in the treatment and control groups conditional on individual's features. To estimate it we may use one of two most popular uplift regression tools: the two model estimator and the response variable transformation based estimator described in [1].
In this contribution we will discuss asymptotic properties of these two approaches and propose another estimator based on the correction of transformed response which combines the advantages of the two existing methods. Experiments on simulated and real data confirm theoretical results and allow us to recommend the proposed new approach.

Keywords: Uplift modeling, linear regression, heterogenous treatment effect
AMS subject classifications: 62J05

Bibliography

[1] K. Rudaś, S. Jaroszewicz (2018) Linear regression for uplift modeling. *Data Mining and Knowledge Discovery*, 32(5), 1275–1305.

Combinatorial regression model in abstract simplicial complexes

Tuesday
September 7th
14:30–14:55

Andrej Srakar[1] **and Miroslav Verbič**[2]

[1]*Institute for Economic Research (IER), Ljubljana*
[2]*School of Economics and Business, University of Ljubljana and Institute for Economic Research (IER), Ljubljana*

Abstract

In regression analysis of the market share data four main parametric type models are prevalent: multinomial logistic regression, attraction models of various types, Dirichlet covariance models, and compositional regression. Nonparametric regressions in simplex spaces include local polynomial (Di Marzio et al, 2015), simplicial spline (Machalová, Hron and Talská, 2019) and simplicial wavelet (Srakar and Fry, 2019) approaches. We extend this arsenal of possibilities with a new type, a completely novel regression perspective, labelled combinatorial regression, based on combining n-tuplets of sampling units into groups and treating them in abstract simplicial complex spaces (Lee, 2011; Korte, Lovász and Schrader, 1991). The novel perspective, estimated in combination with Multivariate Distance Matrix Regression approach (McArdle and Anderson, 2001), allows extensive number of perspectives in the analysis of, for example, triplets, quadruplets or quintuplets (or any n-tuplet) and using as measure of disparity between the units (to construct regressors) different divergence measures. It also allows applications to very small datasets as the number of units in the new model can be expressed in terms of factorial products of units of original sample. We provide the analysis of new approach for different n-tuple combinations using Jensen-Shannon and generalized Jensen-Shannon divergence measures and provide the Gaussian asymptotic limits of the approach with exploring its properties also in a Monte Carlo simulation study. In a short application we present analysis of sessile hard-substrate marine organisms image data from Italian coast areas which allows to explore the new approach in relative abundance data setting.

Keywords: regression models, abstract simplicial complexes, symplectic data, algebraic statistics, algebraic topology
AMS subject classifications: 62M40

Bibliography

[1] Di Marzio, M., Panzera, A. and Venieri, K. (2015). Non-parametric regression for compositional data. *Statistical Modelling.* 15(2), 113–133.
[2] Korte, B., Lovász, L. and Schrader, R. (1991). *Greedoids*, Springer-Verlag, Berlin.

[3] Lee, J.M. (2011). *Introduction to Topological Manifolds*, Springer, New York and London.
[4] Machalová, J., Hron, K. and Talská, R. (2019). Simplicial splines: application and possible extensions. *Proceedings of the 8th International Workshop on Compositional Data Analysis (CoDaWork2019)*. J.J. Egozcue, J. Graffelman and M.I. Ortego (Editors). Universitat Politecnica de Catalunya-BarcelonaTECH.
[5] McArdle, B.H. and Anderson, M.J. (2001). Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology.* 82, 290–297.
[6] Srakar, A. and Fry, T.L.R. (2019). Wavelet regressions for compositional data. *Proceedings of the 8th International Workshop on Compositional Data Analysis (CoDaWork2019)*. J.J. Egozcue, J. Graffelman and M.I. Ortego (Editors). Universitat Politecnica de Catalunya-BarcelonaTECH.

Look-Ahead Screening Rules for the Lasso

Johan Larsson

Department of Statistics, Lund University

Tuesday
September 7th
14:55–15:20

Abstract

The lasso is a popular method for inducing shrinkage and sparsity in the solution vector (coefficients) of regression problems, particularly when the number of predictors far outnumber the number of observations. Solving the lasso for high-dimensional data can, however, be computationally demanding. Fortunately, this computational load can be alleviated via the use of *screening rules*, which screen and discard predictors prior to fitting the model, leading a reduced problem to be solved. Screening rules are particularly effective when fitting a full regularization path: a sequence of models with decreasing penalization. Screening rules can be safe or heuristic. Safe rules certify that discarded predictors are not in the solution; heuristic ones do not. Existing screening rules typically work sequentially or dynamically. Sequential rules screen predictors for the next model along the regularization path, whereas dynamical rules screen during optimization of the current model. There has, however, previously been no attempts to design screening rules that screen further along the path.
In this paper, we present a new screening strategy: *look-ahead* screening rules. Our method uses safe screening rules to find a range of penalty values for which a given predictor cannot enter the model, thereby screening predictors along the remainder of the path. Our screening rules lead to reductions in the time required for screening and also applies to heuristic rules, for which the time required to conduct checks of the optimality conditions to guard against violations of the rules is reduced. In experiments we show that these look-ahead screening rules improve the performance

of existing screening strategies and that the additional cost of screening ahead on the path is marginal.

Keywords: lasso, sparse regression, screening rules, safe screening rules
AMS subject classifications: 62J07

Effectiveness of Indirect Questioning Techniques - Simulation Results

Tuesday
September 7th
17:00–17:25

Pier Francesco Perri[1], Eleni Manoli[2] and Tasos C. Christofides[2]

[1]*Department of Economics, Statistics and Finance "Giovanni Anania", University of Calabria, Italy*
[2]*University of Cyprus, Cyprus*

Abstract

In many fields of applied research, mostly in sociological, economic, demographic and medical studies, misreporting due to untruthful responding represents a non-sampling error that very often occurs especially when survey participants are presented direct questions about sensitive, highly personal or embarrassing issues. Untruthful responses are likely to affect the quality of the collected data and flaw subsequent analyses, including the estimation of salient characteristics of the population under study such as the prevalence of people possessing a sensitive attribute. The problem may be mitigated by adopting indirect questioning techniques which guarantee privacy protection and enhance respondent cooperation.
Making use of direct and indirect questions, we suggest a procedure to detect the presence of liars in sensitive surveys which allows researchers to evaluate the impact of untruthful responses on the estimation of the prevalence of a sensitive attribute. We first introduce the theoretical framework, then apply the proposal to the Warner randomized response method, the unrelated question model, the item count technique, the crosswise model and the triangular model. To assess the effectiveness of the procedure, a simulation study is carried out. Finally, the presence and the amount of liars is discussed in two real studies about racism and mobbing at work. The produced results may encourage survey practitioners to use data-collection modes based on indirect questioning techniques.

Keywords: Direct questioning; randomized response theory; social desirability bias; untruthful responses.
AMS subject classifications: 62P07.

Bibliography

[1] Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys*, Boca Raton: Chapman & Hall/CRC.
[2] Chaudhuri, A., Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys*, Heidelberg: Springer.
[3] Chaudhuri, A., Mukerjee, R. (1988). *Randomized Response: Theory and Techniques*. New York: Marcel Dekker, Inc.
[4] Chaudhuri, A., Christofides, T.C., Rao, C.R. (2016). *Handbook of Statistics 34. Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits*, Amsterdam: Elsevier.
[5] Clark, S.J., Desharnais R.A. (1998). Honest answers to embarrassing questions: Detecting cheating in the Randomized Response Model. *Psychological Methods*, 3, 160–168.
[6] Fox, J.A. (2016). *Randomized Response and Related Methods. Surveying Sensitive Data*, Thousand Oaks: Sage Publication, Inc.

Simple sufficient condition for inadmissibility of Moran's single-split test

Royi Jacobovic[1,2]

[1]*Statistics and Data-Science department, The Hebrew University of Jerusalem*
[2]*Statistics department, University of Haifa*

Tuesday
September 7th
17:25–17:50

Abstract

Suppose that a statistician observes two independent variates X_1 and X_2 having densities $f_i(\cdot;\theta) \equiv f_i(\cdot - \theta)$, $i = 1, 2$, $\theta \in \mathbb{R}$. His purpose is to conduct a test for

$$H : \theta = 0 \quad \text{vs.} \quad K : \theta \in \mathbb{R} \setminus \{0\}$$

with a pre-defined significance level $\alpha \in (0, 1)$. Moran [3] suggested a test which is based on a single split of the data, *i.e.,* to use X_2 in order to conduct a one-sided test in the direction of X_1. Specifically, if b_1 and b_2 are the $(1 - \alpha)$'th and α'th quantiles associated with the distribution of X_2 under H, then Moran's test has a rejection zone

$$(a, \infty) \times (b_1, \infty) \cup (-\infty, a) \times (-\infty, b_2)$$

where $a \in \mathbb{R}$ is a design parameter. Motivated by this issue, the current work includes an analysis of a new notion, *regular admissibility* of tests. It turns out that the theory regarding this kind of admissibility leads to a simple sufficient condition on $f_1(\cdot)$ and $f_2(\cdot)$ under which Moran's test is inadmissible. Furthermore, the same approach leads to a formal proof for the conjecture of DiCiccio [1] addressing that

the multi-dimensional version of Moran's test is inadmissible when the observations are d-dimensional Gaussians.

Keywords: Moran's single-split test, Regular admissibility, Inadmissible test, Data-splitting.
AMS subject classifications: 62C15. 62C07. 62F03.
Acknowledgements: The author would like to thank Ori Davidov for interesting discussions which helped in finding the topic of this work.

Bibliography

[1] DiCiccio, C. J. (2018). *Hypothesis Testing Using Multiple Data Splitting*. Stanford University.
[2] Jacobovic, R. (2021). Simple sufficient condition for inadmissibility of Moran's single-split test. *arXiv:2103.11205*.
[3] Moran, P. A. (1973). Dividing a sample into two parts a statistical dilemma. *Sankhya: The Indian Journal of Statistics*, Series **A**, 329-333.

Generalized Gaussian time series model for EEG data

Tuesday
September 7th
17:50–18:15

N.N. Leonenko[1], Z. Salinger[1], A. Sikorskii[2], N. Šuvak[3] and M.J. Boivin[4]

[1]*School of Mathematics, Cardiff University, Cardiff, United Kingdom*
[2]*Departments of Psychiatry and Statistics and Probability, Michigan State University, East Lansing, Michigan, USA*
[3]*Department of Mathematics, J.J. Strossmayer University of Osijek, Osijek, Croatia*
[4]*Departments of Psychiatry and Neurology and Ophtalmology, Michigan State University, East Lansing, Michigan, USA*

Abstract

Previous analysis of electroencephalogram (EEG) data from [4] showed that stochastic modelling of EEG features can improve the explanation of variation in neurodevelopmental and cognitive outcomes of children who were affected by cerebral malaria. In this analyis, EEG increments were observed as a time series $(X_n, n \in \mathbb{N})$, representing the model for discrete-time observations from the diffusion process $(X_t, t \geq 0)$ with a stationary PDF based on the diffusion process construction presented in [1]. We propose a new strictly stationary time series model with marginal generalized Gaussian distribution and exponentially decaying autocorrelation function for modeling of increments of EEG data collected from Ugandan children during coma from cerebral malaria. The model inherits its appealing properties from the strictly stationary strong mixing Markovian diffusion

with invariant generalized Gaussian distribution (GGD). The GGD parametrization used in this paper was adapted from [3] and comprises some famous light-tailed distributions (e.g., Laplace and Gaussian) and some well known and widely applied heavy-tailed distributions (e.g., Student). Two versions of this model were fit to the data from each EEG channel. In the first model, marginal distributions were from the light-tailed GGD subfamily, and the distribution parameters were estimated using quasi-likelihood approach. In the second model, marginal distributions were heavy-tailed (Student), and the tail index was estimated using the approach based on the emiprical scaling function introduced in [2]. The estimated parameters from models across EEG channels were explored as potential predictors of neurocognitive outcomes of these children 6 months after recovering from illness. Several of these parameters were shown to be importnat predictors even after controlling for nerocognitive scores immediately following cerebral malaria illness and traditional blood and cerebrospinal fluid biomarkers collected during hospitalization.

Keywords: Diffusion discretization, Generalized Gaussian distribution, Tail index, EEG modelling, Elastic net regression.
AMS subject classifications: 37M10, 62M10, 62G07, 62J20, 62P10.
Acknowledgements: The studentship for Z. Salinger is funded through the UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership (project reference 2275322).

Bibliography

[1] Bibby, B. M., Skovgaard, I. M. and Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli.* **11**(2), 191–220.
[2] Grahovac, D., Jia, M., Leonenko, N. and Taufer, E. (2015). Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Statistics.* **49**(6), 1221–1242.
[3] Lutwak, E., Lv, S., Yang, D. and Zhang, G. (2013). Affine moments of a random vector. *IEEE Transactions on Information Theory.* **59**(9), 5592–5599.
[4] Veretennikova, M. A., Sikorskii, A. and Boivin, M. J. (2018). Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria. *Journal of Statistical Distributions and Applications.* **5**(1).

Testing Independence under Biased Sampling

Yaniv Tenzer[1], Micha Mandel[2] and Or Zuk[2]

[1]*Department of Applied Mathematics, Weizmann Institute of Science*
[2]*Department of Statistics, The Hebrew University of Jerusalem*

Tuesday
September 7th
18:15–18:40

Abstract

Testing for dependence between pairs of random variables is a fundamental problem in statistics. In some applications, data are subject to selection bias that can create spurious dependence. An important example is truncation models, in which observed pairs are restricted to a specific subset of the X-Y plane. Standard tests for independence are not suitable in such cases, and alternative tests that take the selection bias into account are required. Here, we generalize the notion of quasi-independence with respect to the sampling mechanism, and study the problem of detecting any deviations from it. We develop two tests statistics motivated by the classic Hoeffding's statistic, and use two approaches to compute their distribution under the null: (i) a bootstrap-based approach, and (ii) a permutation-test with non-uniform probability of permutations. We also handle an important application to the case of censoring with truncation, by estimating the biased sampling mechanism from the data. We prove the validity of the tests, and show, using simulations, that they improve power compared to competing methods for important special cases. The tests are applied to four data sets, two that are subject to truncation, with and without censoring, and two to bias mechanisms related to length bias.

Keywords: quasi-independence, Markov chain Monte Carlo, permutation test, truncation, weighted distribution

Comparison of statistics for testing changed segment in a sample

Wednesday
September 8th
11:00–11:25

Jovita Gudan

Department of Mathematics and Informatics, Vilnius University

Abstract

We propose new test statistics $\mathrm{MR}_{\gamma,n}$ and $T^{ad}_{\gamma,\beta,n}$ for detecting a changed segment in the mean, at unknown dates, in a regularly varying sample model which supports several alternatives of shifts in the mean, including one change point, constant, epidemic and linear form of a change. Our aim is to detect a short length changed segment ℓ^*, assuming ℓ^*/n to be small as the sample size n is large. Statistic $\mathrm{MR}_{\gamma,n}$ is introduced in [1] and is built by taking maximal ratios of weighted moving sums statistics of four subsamples. An important feature of $\mathrm{MR}_{\gamma,n}$ is to be scale free. We obtain the limiting distribution of ratio statistics under the null hypothesis as well as their consistency under the alternative. These results are extended from i.i.d. samples under H_0 to some dependent samples. Meanwhile, $T^{ad}_{\gamma,\beta,n}$ is build by taking maximal ratios of weighted moving sums and its squares statistics without

dividing the sample into four parts. This test statistic is based on the adaptive self-normalized partial sums process [2]. To supplement theoretical results, empirical illustrations are provided by generating samples from symmetrized Pareto and Log-Gamma distributions to compare $\mathrm{MR}_{\gamma,n}$ and $T^{ad}_{\gamma,\beta,n}$ test statistics.

Keywords: change-point detection, changed segment in the mean, epidemic change, Hölder norm statistics, regularly varying random variables.
AMS subject classifications: 62G10, 60F17.

Bibliography

[1] Gudan, J. and Račkauskas, A. and Suquet, Ch. (2021). Testing mean changes by maximal ratio statistics. *Extremes* (submitted).
[2] Račkauskas, A. and Suquet, Ch. (2001). Invariance principles for adaptive self-normalized partial sums processes. *Stochastic Processes and their Applications* 95(1), 63–81.

Semi-Markov Regime Switching Modeling for Anomaly Detection in Time-Series Incidence Data

Emmanouil-Nektarios Kalligeris[1], Alex Karagrigoriou[1], Andreas Makrides[1], Christina Parpoula[2] and Vlad Stefan Barbu[3]

[1] *Lab of Statistics and Data Analysis, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, 83200 Karlovasi, Samos, Greece*
[2] *Department of Psychology, Panteion University of Social and Political Sciences, 17671 Athens, Greece*
[3] *Laboratoire de Mathématiques Raphaël Salem, Université de Rouen-Normandie, UMR 6085, Avenue de l'Université, BP.12, F76801 Saint-Étienne-du-Rouvray, France*

Wednesday
September 8th
11:25–11:50

Abstract

In this work, we transform the typical Markov regime switching model into a semi-Markov regime switching model with the intention of detecting (possibly) hazardous anomalies in time series incidence data. In that context, we consider various underlying distributions for the waiting times between states and we proceed with statistical inference by providing parameter estimates (along with the associated standard errors) and transition probabilities. Finally, examples and applications are provided for illustrative purposes.

Keywords: Anomaly detection, Incidence data, Regime switching, Semi-Markov processes, Time-series.

AMS subject classifications: 62M10, 62F10, 60K15 & 60J05.

Bibliography

[1] V. S. Barbu, A. Karagrigoriou and A. Makrides. *Semi-Markov modelling for multi-state systems*. Methodol. Comput. Appl. Probab., 19:1011–1028, 2017.
[2] N. Limnios and G. Oprişan. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston, 2001.
[3] G. Lindgren. *Markov Regime Models for Mixed Distributions and Switching Regressions*. Scand. J. Stat., 5(2):81–91, 1978.

Predictive Monitoring Using Machine Learning Algorithms and a Real-Life Example on Schizophrenia

Wednesday
September 8th
11:50–12:15

Leo C.E. Huberts[1], Ronald J.M.M. Does[1], Bastian Ravesteijn[2] and Joran Lokkerbol[3]

[1] *University of Amsterdam, Department of Operations Management*
[2] *Erasmus University Rotterdam, School of Economics*
[3] *Trimbos Instititute, Epidemiology*

Abstract

Predictive process monitoring aims to produce early warnings of unwanted events. We consider the use of the machine learning method XGBoost as the forecasting model in predictive monitoring. A tuning algorithm is proposed as the signaling method to produce a required false alarm rate. We demonstrate the procedure using a unique data set on mental health in the Netherlands. The goal of this application is to support healthcare workers in identifying the risk of a mental health crisis in people diagnosed with schizophrenia. The procedure we outline offers promising results and a novel approach to predictive monitoring.

Keywords: Predictive Process Monitoring, Tuning Algorithm, False Alarm Rate, Machine Learning, Mental Health

Monitoring overcrowding in a network of hospital emergency departments

Laura Boyle

Mathematical Sciences Research Centre, Queen's University Belfast, Northern Ireland, United Kingdom

Wednesday
September 8th
12:15–12:40

Abstract

The COVID-19 pandemic has placed a novel strain on health systems internationally and changed the way that patients access medical care. At the start of the COVID-19 pandemic, the number of patients attending Australian emergency departments (EDs) plummeted, with speculation that people were avoiding hospitals [1]. Since then, the number of attendances has been increasing and EDs have been busier than their pre-pandemic operation since January 2021. This research uses near real-time data collected from a publicly available dashboard to monitor and predict the pressure on a network of hospital EDs in Australia. The talk will discuss statistical models for predicting ED overcrowding and outline the challenges associated with causality and missing data.

Keywords: time series, forecasting, health care, missing data.
AMS subject classifications: 62M10,62P99.

Bibliography

[1] Boyle, L.M., Mackay, M., Bean, N. and Roughan-M. (2021). The impact of the COVID-19 Pandemic on South Australia's Emergency Departments – evidence from two lockdowns. *Australian Health Review*.

Sparse structures with LASSO through Principal Components: forecasting GDP components in the short-run

Saulius Jokubaitis, Dmitrij Celov and Remigijus Leipus

Institute of Applied Mathematics, Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius LT-03225, Lithuania

Wednesday
September 8th
13:40–14:05

Abstract

The paper examines the use of sparse methods to forecast the real (in the chain-linked volume sense) expenditure components of the US and EU GDP in the short-run sooner than national statistics institutions officially release the data. We estimate current-quarter nowcasts, along with one- and two-quarter forecasts, by bridging quarterly data with available monthly information announced with a much smaller delay. We solve the high-dimensionality problem of monthly datasets by assuming sparse structures of leading indicators capable of adequately explaining the dynamics of the analyzed data. For variable selection and estimation of the forecasts, we use LASSO together with its recent modifications. We propose an adjustment that combines LASSO cases with principal components analysis to improve the forecasting performance. We evaluated the forecasting performance by conducting pseudo-real-time experiments for gross fixed capital formation, private consumption, imports, and exports over a sample from 2005–2019, compared with benchmark ARMA and factor models. The main results suggest that sparse methods can outperform the benchmarks and identify reasonable subsets of explanatory variables. The proposed combination of LASSO and principal components further improves the forecast accuracy.

Keywords: nowcasting, LASSO, principal components analysis, variable selection, GDP components
AMS subject classifications: 62P20, 62J07
Acknowledgements: The authors would like to thank the anonymous Referee for his/her very constructive and detailed comments and suggestions on the first version of the manuscript. Remigijus Leipus acknowledges the support from the grant No. S-MIP-20-16 from the Research Council of Lithuania.

Statistical methods for modern proteomics

Wednesday
September 8th
14:05–14:30

Mateusz Staniak

University of Wrocław, plac Uniwersytecki 1, 50-137 Wrocław, Poland

Abstract

In this talk, we will present our work on applications of statistical methods to problems in the analysis of mass spectrometry-based proteomics data. Mass spectrometry is a core technology for proteomics. It allows for the identification and quantification of proteins in biological samples. In a mass spectrometry experiment, peptides - smaller fragments of proteins - are ionized, separated based on their mass and charge, and quantified. Resulting data are complex and high-dimensional: they include up to thousands of proteins and tens of thousands of peptides. As the sample

size is small, mass spectrometry data fall into the high dimensional data category ("$p >> n$").
Protein inference is the problem of deriving a list of proteins that are present in the sample based on identified peptides. It is complicated due to three factors: false identifications of peptides based on mass spectra, presence of peptide sequences that can be attributed to multiple proteins (*shared peptides*) and *one-hit wonders* - proteins identified only by a single peptide. Similarly, protein quantification - estimating the relative abundance of proteins in different conditions - is difficult in the presence of shared peptides, as it is not clear how to distribute peptide abundance among their respective proteins. Proteins that are only identified by shared peptides pose a particular challenge.
Thus, the goal of statistical modeling is two-fold: to estimate quantitative profiles at protein-level and to remove both falsely identified proteins, and false assignments between peptides and proteins. We will show how regression analysis based on non-linear mixed models, regularization and robust estimation can solve this problem.

Keywords: applied statistics, biostatistics, multivariate analysis
AMS subject classifications: 62P10, 62J02, 92B15
Acknowledgements: This work is financially supported by the National Science Centre, Poland grant 2020/37/N/ST6/04070.

Bibliography

[1] Nesvizhskii, Alexey I., and Ruedi Aebersold. *Interpretation of shotgun proteomic data.* Molecular & cellular proteomics 4.10 (2005): 1419-1440.
[2] Huang, Ting, et al. *Protein inference: a review.* Briefings in bioinformatics 13.5 (2012): 586-614.

On Estimating the Distribution of the Maximum Likelihood Estimator in Exponential Families

Henning Höllwarth

Technical University of Freiberg

Wednesday
September 8th
14:30–14:55

Abstract

Maximum likelihood (ML) estimation in exponential families, especially for data with complex dependencies such as point patterns, often suffers from an intractable likelihood. As a consequence, we are concerned with theoretical and numerical obstacles. Asymptotic normality of the ML estimator is generally not proven, in fact one may expect a nonstandard asymptotic behavior ([3]). And although an ML

estimate can be obtained by using Markov chain Monte Carlo methods ([4]), the situation is still unsatisfactory due to the computational intensity. Thus, performing a parametric bootstrap to evaluate the uncertainty of an ML estimate may be prohibitively expensive.
We present a simple estimator for the distribution of the ML estimator in exponential families. This estimator uses computationally efficient alternative estimation methods based on pseudo-likelihoods (e. g. [2]) or variational equations (e. g. [1]) and can be easily implemented by using standard software for linear regression. A simulation study illustrates the proposed method for the ML estimator in a Lennard-Jones Gibbs point process model.

Keywords: likelihood inference, parametric bootstrap, Gibbs point process model, Lennard-Jones potential
AMS subject classifications: 62E17, 62F25, 62F40, 62H12, 62M30

Bibliography

[1] Baddeley, A. and Dereudre, D. (2013). *Variational estimators for the parameters of Gibbs point process models*, Bernoulli, 19 (3), 905–930.
[2] Baddeley, A., Coeurjolly, J.-F., Rubak, E. and Waagepetersen, R. (2014). *Logistic regression for spatial Gibbs point processes*, Biometrika, 101 (2), 377–392.
[3] Dereudre, D. and Lavancier, F. (2017). *Consistency of likelihood estimation for Gibbs point processes*, Annals of Statistics, 45 (2), 744–770.
[4] Geyer, C. J. and Thompson, E. (1992). *Constrained Monte Carlo maximum likelihood for dependent data*, Journal of the Royal Statistical Society: Series B (Methodological), 54 (3): 657–683.

Does the Markov Decision Process Fit the Data: Testing for the Markov Property in Sequential Decision Making

Wednesday
September 8th
14:55–15:20

Chengchun Shi

London School of Economics and Political Science

Abstract

The Markov assumption (MA) is fundamental to the empirical validity of reinforcement learning. In this paper, we review the Forward-Backward Learning procedure developed by [1] to test MA in sequential decision making. The test does not assume any parametric form on the joint distribution of the observed data and plays an important role for identifying the optimal policy in high-order Markov decision

processes (MDPs) and partially observable MDPs. It can be naturally coupled with existing reinforcement learning algorithms to improve their performance.

Keywords: Reinforcement Learning; Statistical Inference; Markov Decision Process; Markov Property
AMS subject classifications: 68T05.
Acknowledgements: Shi's research was partially supported by LSE's Research Support Fund in 2021

Bibliography

[1] Shi, C., Wan, R., Song, R., Lu, W., & Leng, L. (2020). Does the Markov decision process fit the data: testing for the Markov property in sequential decision making. In International Conference on Machine Learning (pp. 8807-8817). PMLR.

Distributions transformations and Stein's method

Nikolai Slepov

Lomonosov Moscow State University

Wednesday
September 8th
17:00–17:25

Abstract

The Stein method is a powerful tool permitting to estimate the accuracy of various approximations for a target probability distribution by other ones. On this way a number of important results concerning the convergence rates in classical limit theorems were derived. Moreover, in certain cases the sharp bounds of closeness (in a sense) of laws under consideration were obtained. Distributions transformations play an essential role in development of this method. We mention, e.g., the zero-bias transformation ([2]). It is well-known that after the Stein fundamental work on Gaussian distribution, this method was modified to comprise the Poisson, gamma, geometric, negative-binomial and other laws, see also a recent book [1]. The choice of appropriate metric for a studied problem is also important. In this regard we refer, e.g., to [6] where the optimal bounds for normal approximation were found involving the zeta-metrics.
Our goal is the Stein method development for investigating the distributions of random sums of random variables. We employ centered equilibrium transformation for approximations by the Laplace law (see [3, 5]) and use the generalized equilibrium transform (see [4]) for geometric approximation. In both cases new upper bounds in limit theorems for random sums are established. An analogue of the famous Rényi theorem is provided with exact bound. There are numerous applications of Stein's

techniques, e.g. in geometry, analysis of DNA sequences, random graphs among others. We concentrate on some applications to queueing theory.

Keywords: zeta-metrics, equilibrium transform, Rényi theorem.
AMS subject classifications: 60F05
Acknowledgements: The author is grateful to Prof. A. V. Bulinski for useful discussions.

Bibliography

[1] Arras, B. and Houdré, C. (2019) *On Stein's method for infinitely divisible laws with finite first moment*, Springer International Publishing.
[2] Goldstein, L. and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. of Appl. Probab.* 7(4), 935-952.
[3] Pike, J. and Ren, H. (2014). Stein's Method and the Laplace Distribution. *ALEA*, 11(2), 571-587.
[4] Shevtsova, I. and Tselishchev, M. (2020). A Generalized Equilibrium Transform with Application to Error Bounds in the Renyi Theorem with No Support Constraints. *Mathematics*, 8, 577.
[5] Slepov, N. (2021). The convergence rate of random geometric sum distribution to the Laplace law. *Theory Probab. Appl.*, 66(1), 149-174.
[6] Tyurin, I. (2011). On the convergence rate in Lyapunov's theorem. *Theory Probab. Appl.*, 55(2), 253-270.

An analogue of the Feynman-Kac formula for higher order evolution equations

Wednesday
September 8th
17:25–17:50

Mariia Platonova

St. Petersburg Department of V.A. Steklov Institute; Chebyshev Laboratory, St. Petersburg State University, St. Petersburg, Russian Federation

Abstract

It is well known that a solution to the Cauchy problem for the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + V(x)u,\ u(0,x) = \varphi(x)$$

can be represented as the expectation of a functional of a Wiener process. Namely,

$$u(t,x) = \mathbf{E}\,\varphi(x - w(t)) \exp\Big(\int_0^t V(x - w(s))ds\Big), \tag{1}$$

where $w(t)$ is a standard Wiener process. The formula (1) is called the Feynman-Kac formula (see [2], p. 308). If the differential operator in the evolution equation is of order $2m > 2$ and has the form

$$\frac{\partial u}{\partial t} = \frac{(-1)^{m+1}}{(2m)!}\frac{\partial^{2m} u}{\partial x^{2m}} + V(x)u, \tag{2}$$

then any representation of the Cauchy problem solution analogous to (1) with $w(t)$ replaced by some random process is impossible, since the fundamental solution of (2) is not a probability measure. However, in the paper [1], a probabilistic method was proposed for constructing an approximation of the Cauchy problem solution for (2) in the case $V = 0$ by mean values of functionals of stochastic processes. Using ideas and methods of [1], we are constructing an approximation of the Cauchy problem solution for the evolution equation (2). The approximating operators take the form of expectations of functionals of a certain random point field.

Keywords: Evolution equations, Poisson random measures, Feynman-Kac formula.
AMS subject classifications: 28C20, 35K25, 60G55.
Acknowledgements: This research is supported by the Russian Science Foundation grant 19-71-30002.

Bibliography

[1] Platonova, M.V. (2018). A probabilistic representation of the Cauchy problem solution for an evolution equation with the differential operator of the order greater than 2. *Journal of Mathematical Sciences* 229:6, 744–755.
[2] Reed, M.C., Simon, B. (1975). *Methods of Modern Mathematical Physics*, vol. 2, Academic, New York.

A CLT for degenerate diffusions with periodic coefficients, and application to homogenization of linear PDEs

N. Sandrić and I. Valentić

University of Zagreb, Faculty of Natural Sciences

Wednesday
September 8th
17:50–18:15

Abstract

Let $\mathcal{L}^\varepsilon$, $\varepsilon > 0$, be a second-order elliptic differential operator of the form

$$\mathcal{L}^\varepsilon \;=\; \big(a(\cdot/\varepsilon) + \varepsilon^{-1} b(\cdot/\varepsilon)\big)^{\mathrm{T}} \nabla + 2^{-1}\mathrm{Tr}\big(c(\cdot/\varepsilon)\,\nabla\nabla^{\mathrm{T}}\big)\,.$$

We discuss periodic homogenization (that is, asymptotic behavior of the solution as $\varepsilon \to 0$) of the associated elliptic boundary-value problem as well as the parabolic initial-value problem in the case of degenerate (possibly vanishing on a set of positive Lebesgue measure) diffusion coefficient $c(x)$.

Our approach to this problem relies on probabilistic techniques, ideas for which go back to M. I. Freĭdlin [2]. We first show that the (appropriately centered) diffusion process associated to $\mathcal{L}^\varepsilon$ satisfies a functional CLT with Brownian limit as $\varepsilon \to 0$, and then by employing probabilistic representation (the Feynman-Kac formula) of the (viscosity) solutions to the elliptic boundary-value and the parabolic initial-value problem (see [3]) we conclude the homogenization result. In the non-degenerate (uniformly elliptic) case these steps can be carried out by combining classical PDE results (existence of a smooth solution to the corresponding Poisson equation) and the fact that the underlying diffusion process does not show a singular behavior in its motion, that is, it is irreducible. In the case of a degenerate diffusion part, this deficiency is compensated by the assumption that the underlying diffusion process with positive probability reaches the part of the state space where the diffusion term is non-degenerate. In other words, this condition ensures irreducibility of the process. Also, in this case it is not clear that we can rely on PDE techniques therefore the analysis of a solution to the corresponding Poisson equation is completely based on stochastic analysis tools.

In the closely related article [4], by also employing probabilistic methods, the authors are concerned with the same questions, however, unfortunately, there seems to be a doubt about their proof of the functional CLT. Under slightly weaker assumptions (and by employing different techniques) we resolve this issue, or at least suggest an alternative approach to the problem.

Keywords: Periodic homogenization, Degenerate diffusion processes, CLT, Feynman-Kac formula

AMS subject classifications: 60H30

Bibliography

[1] N. Sandrić. and I. Valentić. *A CLT for degenerate diffusions with periodic coefficients, and application to homogenization of linear PDEs*. to appear in J. Differ. Equ. 30 pp (2021).

[2] M.I. Freĭdlin. *Dirichlet problem for an equation with periodic coefficients depending on a small parameter*. Teor. Verojatnost. i Primenen. 9:133–139. (1964).

[3] E. Pardoux and A. Rascanu. *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*. Springer International Publishing. (2014).

[4] M. Hairer and E. Pardoux. *Homogenization of periodic linear degenerate PDEs*. J. Funct. Anal. 255(9):2462–2487. (2008).

Penalized model-based clustering for three-way data structures

Andrea Cappozzo[1], Alessandro Casa[2] and Michael Fop[2]

[1]*Department of Mathematics, Politecnico di Milano*
[2]*School of Mathematics & Statistics, University College Dublin*

Thursday
September 9th
11:00–11:25

Abstract

Recently, there has been an increasing interest in developing statistical methods able to find groups in matrix-valued data [1]. To this extent, Matrix Gaussian Mixture Models (MGMM) provide a natural extension to the popular model-based clustering based on Normal mixtures [2]. Unfortunately, the overparametrization issue, already affecting the vector-variate framework, is further exacerbated when it comes to MGMM, since the number of parameters scales quadratically with both row and column dimensions. In order to overcome this limitation, the present work introduces a sparse model-based clustering approach for three-way data structures. By means of penalized estimation, our methodology shrinks the estimates towards zero, achieving more stable and parsimonious clustering in high dimensional scenarios. Our contribution is the direct extension to the matrix-variate framework of the penalized model-based clustering developed in [3]. An application to satellite images underlines the benefits of the proposed method.

Keywords: Model based clustering, Matrix distribution, EM-algorithm, Penalized likelihood, Sparse matrix estimation
AMS subject classifications: 62H30, 62H35

Bibliography

[1] Sarkar, S., Zhu, X., Melnykov, V. and Ingrassia, S. (2020). On parsimonious models for modeling matrix data. *Comp. Stat. and Data Anal.* 142, 106822.
[2] Viroli, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Stat. and Comp.* 21(4), 511–522.
[3] Zhou, H.,Pan, W. and Shen, X. (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electr. J. of Stat.* 3, 1473–1496.

Protecting Classifiers from Attacks

Roi Naveiro[1], Víctor Gallego[1], Alberto Redondo[1], David Ríos Insua[1] and Fabrizio Ruggeri[2]

[1]*Institute of Mathematical Sciences (ICMAT-CSIC)*
[2]*CNR-IMATI*

Thursday
September 9th
11:25–11:50

Abstract

In multiple domains such as malware detection, automated driving systems, or fraud detection, statistical classification algorithms are susceptible of being attacked by malicious agents who are able to perturb the value of the covariates of instances to attain certain goals. Such problems pertain to the field of adversarial machine learning and have been dealt with mostly through game-theoretic ideas with strong underlying common knowledge assumptions. These are not realistic in numerous application domains. We present an alternative statistical framework that accounts for the lack of knowledge about the attacker's behavior using adversarial risk analysis. A key ingredient is the ability to sample from the distribution of originating covariates $\mathbf{x}$ given the possibly attacked observed ones $\mathbf{x}'$. We propose a sampling procedure based on approximate Bayesian computation (ABC, [3]). Within it, we simulate the attacker's problem taking into account our uncertainty about his elements.

The basic idea of our approach is that, when observing the possibly modified covariates $\mathbf{x}'$, we must model our uncertainty about the latent originating ones $\mathbf{x}$ through a distribution $p(\mathbf{x}|\mathbf{x}')$. Upon observing $\mathbf{x}'$, inference about the latent origins $\mathbf{x}$ given the observed covariates must be undertaken. This entails estimating $p(\mathbf{x}|\mathbf{x}')$ or, at least, sampling from it. To do so, we need to define an *attacking model* $p(\mathbf{x}'|\mathbf{x})$, that is, a model of our beliefs about how the attacker will modify covariates $\mathbf{x}$, that reflects the existing lack of knowledge about the adversary's elements. From a modeling perspective this is hard as it requires strategic thinking about the adversary. Stemming from the work in [1], we propose a formal Bayesian decision theoretic approach to sample from $p(\mathbf{x}'|\mathbf{x})$, based in Adversarial Risk Analysis (ARA) [2]. Once samples $\mathbf{x}' \sim \mathbf{p}(\mathbf{x}'|\mathbf{x})$ are available, an ABC based scheme is proposed to generate samples from $p(\mathbf{x}|\mathbf{x}')$. Finally, upon observing instance $\mathbf{x}'$, a classification decision is made based on an estimate of the probability of the class given the observed covariates $p(y|\mathbf{x}')$, which is computed marginalizing out every possible originating instance $\mathbf{x}$.

Keywords: Statistical Classification, Adversarial Risk Analysis, Approximate Bayesian Computation.

AMS subject classifications: 62H30, 91A35.

Acknowledgements: This work was partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute and a BBVA Foundation project.

Bibliography

[1] Naveiro, R., Redondo, A., Insua, D. R., & Ruggeri, F. (2019). *Adversarial classification: An adversarial risk analysis approach.* International Journal of Approximate Reasoning, 113, 133-148.

[2] Rios Insua, D., Ríos, J., & Banks, D. (2009). *Adversarial risk analysis.* Journal of the American Statistical Association, 104(486), 841-854.

[3] Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). *Approximate Bayesian computation (ABC) in practice.* Trends in ecology & evolution, 25(7), 410-418.

Supervised classification of replicated point patterns

Kateřina Koňasová and Jiří Dvořák

Faculty of Mathematics and Physics, Charles University, Czech Republic

Thursday
September 9th
11:50–12:15

Abstract

Spatial point processes are mathematical models describing the arrangement of objects randomly placed in space. Such models are of particular interest in many scientific disciplines, including biology, ecology, or material science [2]. Supervised classification is one of the fundamental problems in statistics and machine learning. For replicated point patterns, the task is to predict the group membership for a newly observed pattern using the information from a training collection of patterns with known labels. Standard classification techniques such as the k-nearest neighbours algorithm or the kernel regression method can be employed, provided that the dissimilarities [3] between any pair of patterns can be quantified.

We propose a general classification method that can be used for both stationary and non-stationary point processes in d-dimensional Euclidean space and can be further generalized to more complicated settings. Special attention is paid to dissimilarity measures based on point process functional summary characteristics, e.g. the pair correlation function. Various characteristics can be employed, depending on the preliminary knowledge about the investigated point patterns. Thus, it is possible to adapt the dissimilarity measure precisely to the problem at hand. Furthermore, a link between methods for point patterns and functional data [1] can be established. Instead of comparing the patterns directly, one can compare the values of a selected functional characteristic estimated from the individual patterns.

We illustrate the performance of the kernel regression classifier with a dissimilarity measure based on point process functional characteristics via a set of simulation experiments oriented towards the influence of different attributes of the classification procedure. Results of these experiments indicate that the proposed method is successful in solving the problem.

Keywords: Spatial point patterns, Supervised classification, Kernel regression, Dissimilarity measure
AMS subject classifications: 60G55, 62H30
Acknowledgements: This work has been supported by The Charles University Grant Agency, project no. 1198120, and The Czech Science Foundation, project no. 19-04412S.

Bibliography

[1] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis Theory and practice*,
Springer-Verlag, New York.

[2] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2004). *Statistical Analysis and Modelling of Spatial Point Patterns* John Wiley and Sons, Chichester.
[3] Mateu, J., Schoenberg, F. P., Diez, D. M., González, J. A. and Lu, W. (2015). On measures of dissimilarity between point patterns: classification based on prototypes and multidimensional scaling. *Biometrical Journal* 57, 340–358.

Analysis of Spatio-Temporal Data Using Bayesian and Formal Methods

Thursday
September 9th
12:15–12:40

Laura Vana

Vienna University of Economics and Business

Abstract

Urban mobility applications are characterized by highly dynamic spatial and temporal patterns. In this paper we present a case study for the City of Milan, Italy, where we use an aggregated measure of crowdedness derived from mobile phone traffic data for a high-resolution grid over 10 min time intervals. We build a spatio-temporal Bayesian model which is able to capture spatio-temporal dependence in crowdedness, while also identifying areas which share similar temporal behavior. The Bayesian paradigm offers the possibility of simulating future spatio-temporal trajectories of crowdedness measure from the posterior predictive distribution while coherently accounting for uncertainty in the model parameters. Formal methods are employed to formulate spatio-temporal properties for the system such as the presence of temporary overload or that of a fault tolerance and the probability of satisfaction of these properties is obtained through statistical model checking. Results provide a deeper understanding of urban dynamic in Milan, as well as exemplify the value added in combining Bayesian modeling with formal verification methods.

Keywords: Bayesian spatio-temporal model, mobile phone data, population mobility, spatio-temporal clustering, statistical model checking
AMS subject classifications: 62P12
Acknowledgements: This is based on joint work with Annalisa Cadonna, Gregor Kastner (Alpen-Adria University Klagenfurt), Laura Nenzi (TU Wien) and Ennio Visconti (TU Wien). The author acknowledges funding from the Austrian Science Fund (FWF) for the project "High-dimensional statistical learning: New methods to advance economic and sustainability policies" (ZK 35), jointly carried out by WU Vienna University of Economics and Business, Paris Lodron University Salzburg, TU Wien, Alpen-Adria University Klagenfurt and the Austrian Institute of Economic Research (WIFO).

Quantifying and Assessing Climate Resilience of Plant-Soil Systems Through Response Diversity

Janne Kaseva

Natural Resources Institute Finland, Tietotie 4, Jokioinen FI-31600, Finland

Thursday
September 9th
13:40–14:05

Abstract

Weather variability is responsible for a significant portion of global yield variability. However, soil is an increasingly essential component in modeling the effects of climate change and limited empirical research has been conducted clarifying the relationship between climate, soils, and crop yields. This study focuses on revealing the importance of addressing the shifts in climate, soil, and crop combinations. This approach based on response diversity could allow a more valid assessment of diversity in terms of the response to climate variability and change. In the case of added value by response diversity, this approach could provide a generic procedure as a practical tool for resilience management.

Long-term data on wheat cultivars were collected from nine European countries [1]. Linear mixed models were used to analyse soil- and climate-based yield estimates for chosen agro-climatic variables. Multivariate methods such as principal component analysis and cluster analysis were used to classify weather variables and cultivars based on their yield responses to weather. Finally, cultivar-based diversity indices were formed for each country, enabling comparison of true diversities as measured by the exponential of the Shannon indices.

The proposed approach for the empirical identification of response diversity to manage resilience and adaptive capacity to global change creates added value by guiding tailored diversification. Once the key diversity that fosters resilience has been identified, more resilience can be achieved with less diversity. This approach is described in more detail in Kaseva et al. [2], and could be used in many areas of research, such as supply chains.

Keywords: Climate resilience, Response Diversity, Linear Mixed Models, PCA-Clustering
AMS subject classifications: 92B06

.

Bibliography

[1] Kahiluoto, H., Kaseva, J., Balek, J., Olesen, J., Ruiz-Ramos, M., Gobin, A., Kersebaum, KC., Takác, J., Ruget, F., Ferrise, R., Bezak, P., Capellades, G., Dibarik, C., Mäkinen, H., Nendel, C., Ventrella, D., Rodríguez, A., Bindi, M., Trnka, M. (2019). Decline in Climate Resilience of European Wheat. Proceedings of the National Academy of Sciences of the United States of America (PNAS) 116 1: 123-128.

[2] Kaseva, J., Himanen, S., Kahiluoto, H. (2019). Managing Diversity for Food System Resilience. Advances in Food Security and Sustainability 4.

Classification Based on Multivariate Mixed Type Longitudinal Data with an application to the EU-SILC database

Thursday
September 9th
14:05–14:30

Jan Vávra and Arnošt Komárek

Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Abstract

Many nowadays studies gather data of diverse nature (numeric quantities, binary indicators or ordered categories) on the same units repeatedly over time. We present a statistical model capable of joint modelling of several so called *mixed-type* longitudinal outcomes, which also accounts for possible dependencies among investigated outcomes. A thresholding approach to link binary or ordinal variables to their latent numeric counterparts allows us to jointly model all, including latent, numeric outcomes using a multivariate version of the linear mixed-effects model [4]. We avoid the independence assumption over outcomes by relaxing the variance matrix of random effects to a completely general positive definite matrix. Moreover, we follow Model Based Clustering (MBC, [1]) methodology to create a mixture of such models to model heterogeneity in temporal evolution of considered outcomes. The estimation of such hierarchical model is approached by Bayesian principles with the use of Markov Chain Monte Carlo (MCMC, [2]) methods.
The ability to discover different patterns is demonstrated on the EU-SILC [3] dataset consisting of Czech households each followed for four years in a time span 2005 – 2018. Households were classified into groups with similar evolution of several closely related indicators of monetary poverty based on estimated classification probabilities.

Keywords: Multivariate longitudinal data, Mixed type outcome, Model based clustering, Classification, EU-SILC
AMS subject classifications: 62H30
Acknowledgements: This research was supported by the Czech Science Foundation (GAČR) grant 19-00015S.

Bibliography

[1] Banfield, J., D. and Raftery, A., E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics* 49(3), 803–821.

[2] Brooks, S. and Gelman, A. and Jones, G. and Meng, X. (2011). *Handbook for Markov chain Monte Carlo*, Taylor & Francis
[3] European Union Statistics on Income and Living Conditions. URL https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions
[4] Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics* 38(4), 963–974.

Clustering multivariate functional data using unsupervised binary trees

Steven Golovkine[1], **Nicolas Klutchnikoff**[2] **and Valentin Patilea**[3]

[1]*Groupe Renault & CREST - UMR 9194, Rennes, France*
[2]*Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France*
[3]*Ensai, CREST - UMR 9194, Rennes, France*

Thursday
September 9th
14:30–14:55

Abstract

We propose a model-based clustering algorithm for a general class of functional data for which the components could be curves or images. The random functional data realizations could be measured with error at discrete, and possibly random, points in the definition domain. Based on [1], the idea is to build a set of binary trees by recursive splitting of the observations. At each node of the tree, a model selection test is performed, after expanding the multivariate functional data into a well chosen basis. We consider the Multivariate Functional Principal Component basis, developed in [2]. Similarly to [3], using the Bayesian Information Criterion, we test whether there is evidence that the data structure is a mixture model or not at each node of the tree. The number of groups are determined in a data-driven way and does not have to be pre-specified before the construction of the tree. Moreover, the tree structure allows us to consider only a small number of basis functions at each node. The new algorithm provides easily interpretable results and fast predictions for online data sets. Results on simulated datasets reveal good performance in various complex settings. The methodology is applied to the analysis of vehicle trajectories on a German roundabout. The open-source implementation of the algorithm can be accessed at https://github.com/StevenGolovkine/FDApy. Complete version of the work is available at arxiv:2012.05973.

Keywords: Gaussian mixtures, Model-based clustering, Multivariate Functional Principal Components
AMS subject classifications: 62R10

Acknowledgements: The authors wish to thank Groupe Renault and the ANRT for their financial support via the CIFRE convention no. 2017/1116.

Bibliography

[1] Fraiman, R., Ghattas, B. and Svarc, M. (2013). Interpretable clustering using unsupervised binary trees. *Advances in Data Analysis and Classification*, 7.
[2] Happ, C. and Greven, S. (2018). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association* 113 649–659.
[3] Pelleg, D. and Moore, A. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *In Proceedings of the 17th International Conf. on Machine Learning* 727–734.

Goodness of Lift in Collision Insurance

Thursday
September 9th
14:55–15:20

Yves Staudt[1], **Julien Trufin**[2] **and Joël Wagner**[3]

[1] *University of Applied Sciences of the Grisons*
[2] *Université libre de Bruxelles*
[3] *University of Lausanne*

Abstract

For calculating non-life insurance premiums actuaries traditionally rely on seperate severity and frequency models using covariates to explain the claims loss exposure. The aim of this paper is to model the pure premium as a combination of the severity and frequency models. The frequency and severity models are calibrated with the help of two model families, the gradient boosting models (GBM) and generalized additive models (GAM). GBM are optimized with the help of the exponential family deviance as a loss function, i.e. the Gamma deviance in the case of the severity and with the help of the traditional root mean squared error. Their performance is measured with the help of goodness-of-fit and goodness-of-lift statistics. The goodness-of-lift statistics, such as the area between Lorenz and concentration curves, allows to measure the accuracy of the predictions in relation to the customer past expenses. In our application, we rely on a dataset covering the loss exposure of a Swiss collision insurance portfolio covering the period from 2011 to 2015.

Keywords: gradient boosting models, machine learning, performance analysis, goodness-of-lift, goodness-of-statistics.
AMS subject classifications: 62G08,62G35.

Bibliography

[1] Yves Staudt and Jöel Wagner (2021), "Assessing the Performance of Random Forests for Modeling Claim Severity in Collision Car Insurance. " *Risks*, vol. 9(3).
[2] Roel Henckaerts, Marie-Pier Côté, Katrien Antonio and Roel Verbelen (2021), "Boosting Insights in Insurance Tariff Plans with Tree-Based Machine Learning Methods." *North American Actuarial Journal*, vol. **0**, pp. 1-31.
[3] Michel Denuit, Dominik Sznajder and Julien Trufin (2019), "Model selection based on Lorenz and concentration curves, Gini Indices and convex order." *Insurance: Mathematics and Economics*, vol. **89**, pp. 128-139.
[4] Yves Staudt and Jöel Wagner (2020), "Comparison of machine learning and traditional severity-frequency regression models for car insurance pricing. " *working paper, University of Lausanne.*
[5] Roel Henckaerts, Katrien Antonio, Maxime Clijsters and Roel Verbelen (2018), "A data driven binning strategy for the construction of insurance tariff classes." *Scandinavian Actuarial Journal*, vol. **1238**, pp. 1-25.

Yield curve modelling in insurance

Matúš Padyšák

Comenius University, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics

Thursday
September 9th
17:00–17:25

Abstract

Yield curves are an essential part of financial markets. In practice, the most common are parsimonious parametric models and spline-based models. The paper is focused on the application of yield curves in the insurance and to regulations of Solvency II (see [1]). Legislative programme Solvency II requires for a yield curve to be fitted on liquid instruments until the last liquid point, from which the curve is extrapolated to the Ultimate Forward Rate for a long pre-determined horizon. The forward curve has to converge to the Ultimate Forward Rate with a pre-determined accuracy of one basis point. The European Insurance and Occupational Pensions Authority recommends using the spline-based Smith-Wilson model in the insurance applications. The Ultimate Forward Rate is one of the Smith-Wilson parameters, and the convergence is easily ensured by controlling for the speed of convergence. This paper studies whether it is possible to require the same conditions of convergence for the parametric models such as Svensson or the Five-Factor model. This paper aims to present several calibration methods and methodology for the Ultimate Forward Rate convergence in parametric models. Finally, cross-validation techniques and further statistical tests are suggested for accuracy testing.

Keywords: yield curve, Ultimate Forward Rate, calibration, testing

AMS subject classifications: 91B30, 91G30
Acknowledgements: This research was supported by the Slovak grant VEGA 2/0096/21.

Bibliography

[1] EIOPA (2019). *Technical documentation of the methodology to derive EIOPA's risk-free interest rate term structures*, Published online https://www.eiopa.europa.eu/sites/default/files/risk_free_interest_rate/12092019-technical_documentation.pdf

The Partial Least Squares-Path Modeling for the building of Social Composite Indicators

Rosanna Cataldo

Department of Social Sciences of University of Naples "Federico II", Naples, Italy

Thursday
September 9th
17:25–17:50

Abstract

Many social phenomena are complex and therefore difficult to measure and to evaluate. A phenomenon is defined as complex when the relevant aspects of a particular problem cannot be captured by using a single perspective. It is necessary to consider the concept formed by different dimensions, each representing different aspects of it, which interact with each other. For this reason, most of the time, the complexity implies also multidimensionality. The goal of much research in social, economic and political fields is to obtain a whole description of the various facets of this complex phenomenon, through a suitable synthesis of the associated elementary indicators. Research, in the last years, has been focusing on the development and use of a system of composite indicators in order to obtain a global description of a complex phenomenon and to convey a suitable synthesis of information. The existing literature offers several alternative methods for obtaining a composite indicators. In this context my research line is placed. In detail, my works focus on building of composite indicators system through to Structural Equation Modeling (SEM), specifically with the use of Partial Least Squares-Path Modeling (PLS-PM), which allows you to estimate causal relationships, defined according to a theoretical model linking two or more latent complex concepts, each measured through a number of observable indicators. The aim of the work is to demonstrate how PLS-PM could help you to build composite indicators system, in order to provide a better measure of more complex social phenomena. To illustrate the importance of the PLS-PM, a social composite indicator will be described.

Keywords: Composite Indicators, PLS-PM, Higher-order Constructs
AMS subject classifications: 62P25
Bibliography

[1] Cataldo, R., Grassia, M.G., Lauro, N.C., Marino, M. (2017). Developments in Higher-Order PLS-PM for the building of a system of Composite Indicators *Quality & Quantity*, Springer, 51(2), 657–674
[2] Chin, W. W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly*, 22(1), vii - xvi
[3] Chin, W.W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. Marcoulides G. A. (Ed): *Modern Business Research Methods*, Mahwah, NJ: Lawrence Erlbaum Associates, 295 - 336
[4] Lauro, C.N., Grassia, M.G.,Cataldo, R. (2018). Model based composite indicators: New developments in partial least squares-path modeling for the building of different types of composite indicators *Social Indicators Research*, Springer, 135(2), 421- 455
[5] Lohmöller, J. B. (1989). Latent Variable Path Modeling with Partial Least Squares. *Physica*, Verlag, Heidelberg, Germany
[6] Saisana, M., Tarantola, S. (2002). State-of-the-art. Report on Current Methodologies and Practices for Composite Indicator Development. *Institute for the Protection and Security of the Citizen Econometrics and Statistical Support to Antifraud Unit*
[7] Tenenhaus, M., Esposito, Vinzi, V., Chatelin, Y. M., and Lauro, N. C. (2005). PLS Path Modeling. *Computational Statistics and Data Analysis*, 48(1), 159 - 205

Modeling the superposition of dependent binary signals with hidden Markov models

Laura Jula Vanegas

Institute for Mathematical Stochastics, University of Göttingen, Germany

Thursday
September 9th
17:50–18:15

Abstract

Hidden Markov models (HMMs) are a widely used tool for the modeling of non observable Markov chains, for example to model current recordings of ion channels in the cell. Recently, there is a growing interest in the multivariate case, namely for multidimensional Markov chains $(X_k)_{k\in\mathbb{N}}$, $X_k = (X_k^{(1)}, \dots, X_k^{(\ell)})$. Of course, if the entries of X_k are independent, the problem can be solved with standard tools. However, the situation gets more complex once independence can not be assumed and the complete multidimensional vector is not observable. This becomes even harder when only the superposition of signals can be detected, namely,

$$S_k = \sum_{i=1}^{\ell} X_k^{(\ell)}$$

for $k \in \mathbb{N}$. This is the case in our application, were only the total measure of a piece of membrane with multiple ion channels can be measured, and there is evidence that the channels may not be independent.

In this talk we concern ourselves with a fixed dimension ℓ, binary chains $(X_k^{(\ell)})_{k\in\mathbb{N}}$ and a HMM structure on top of the superposition process $(S_k)_{k\in\mathbb{N}}$. Surprisingly, the only model in the literature to tackle this problem is the Chung-Kennedy model [1], with which we compare. We model the multidimensional chain's transition matrix with a vector norm dependent Markov chain model (VND) and provide an easy characterization. This model ensure us that we can recover the full dependency behavior from the superposition process, and model a wide range of dependency behaviors. Moreover, within this model the sum is again a Markov chain, and therefore, we can use a modified Baum-Welch algorithm in order to estimate the parameters from the HMM. More details can be found in [2].

Keywords: Hidden Markov models, vector norm dependency, permutation invariance, lumping property, aggregated data.
AMS subject classifications: 62M05, 62P10, 62H12
Acknowledgements: This work was supported by the DFG projects SFB803 and the RTG2088.

Bibliography

[1] Chung S.H., Kennedy R. A. (1996). Coupled Markov chain model: Characterization of membrane channel currents with multiple conductance sublevels as partially coupled elementary pores. *Math. Biosci.* 133(2), 111–137.
[2] Vanegas L.J., Eltzner B., Rudolf D., Dura M., Lehnart S. E., Munk, A. (2021). Analyzing cross-talk between superimposed signals: Vector norm dependent hidden Markov models and applications *Submitted*. https://arxiv.org/abs/2103.06071

Stochastic modelling of the bactericidal potency

Thursday
September 9th
18:15–18:40

Péter Kevei[1], Máté Szalai[2] and Dezső Virok[3]

[1]*Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, 6720 Szeged, Hungary; e-mail:* `kevei@math.u-szeged.hu`
[2]*Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, 6720 Szeged, Hungary; e-mail:* `szalaim@math.u-szeged.hu`
[3]*Department of Medical Microbiology and Immunobiology, University of Szeged, Dóm tér 10, 6720 Szeged, Hungary; e-mail:* `virok.dezso.peter@med.u-szeged.hu`

Abstract

Branching processes are classical tools to model cell proliferation. However, for estimation of bactericidal potency of antibiotics only deterministic models were used. We develop a stochastic model for bacterial growth, where the size of the bacterial population follows a Galton–Watson branching process, and the offspring distribution depends on the concentration of the antibiotic.
For a fixed antibiotic concentration $c > 0$ we provide an estimator for the offspring mean $m(c)$, and show that the estimator is weakly consistent and asymptotically normal. Assuming that the offspring mean has the form $m(c) = 2/(1 + \alpha c^{\beta})$ with a suitable $\alpha > 0$, $\beta > 0$, we obtain an estimator for the parameters (α, β) and investigate its properties. We provide a consistent estimator for the minimal inhibitory concentration (MIC), which is the smallest concentration preventing bacterial growth, an important parameter in (micro)biology.
As a real data we used measurements of *Chlamydia trachomatis* growth which was analyzed by a novel quantitative PCR method treated by 4 different antibiotics at 12 different antibiotic concentrations. We found that our model fits very well to this data.
The process is also simulated as follows. We take α and β values, and c_i antibiotic concentrations. In addition of these values, we calculate the expected values of $m(c_i)$ from the formula, from which we obtain the vectors $\mathbf{p}^{(i)} = (p_0^{(i)}, p_1^{(i)}, p_2^{(i)}) = (1 - \frac{m(c_i)}{2}, 0, \frac{m(c_i)}{2})$. The whole process is simulated according to these progeny distributions. In the case of simulation, the MIC value is determined similary.

Keywords: Galton–Watson branching process, stochastic modelling, MIC, simulation
AMS subject classifications: 60J80

On the stochastic orders of extremes order statistics

Luigi-Ionut Catana

Faculty of Mathematics and Computer Science, Mathematical Doctoral School, University of Bucharest, Str. Academiei nr. 14, sector 1, C.P. 010014, Bucharest, Romania; luigi_catana@yahoo.com

Friday
September 10th
11:00–11:25

Abstract

In this paper we prove that we can have hazard rate and reversed hazard rate orders of extremes order statistics in the case of quadratic transmuted distributions for a family parameters, although the parameters are not comparable in some sense.

Keywords: Stochastic order, Transmuted distribution, Order statistics
AMS subject classifications: 60E15

Posterior properties of the Weibull model for record data

Friday
September 10th
11:25–11:50

Zoran Vidović

Teacher Education Faculty, University of Belgrade, Belgrade 11000, Serbia

Abstract

The flexibility of the Weibull distribution makes it a preferable model in various reliability studies. In Bayesian framework, when only the sample of record values is available, we discuss necessary and sufficient conditions on the improper priors that lead to proper posteriors of model parameters. The finiteness of the posterior moments is also examined. Obtained results are illustrated using different objective priors. These results extend some previous work on this topic.

Keywords: Weibull distribution, objective prior, proper posterior, records.
AMS subject classifications: 62F15, 62N05

Bibliography

[1] Ramos, E., Ramos, P. L. and Louzada, F. (2020). *Posterior properties of the Weibull distribution for censored data*, Statistics and Probability Letters, 166, 108873.

Testing in Interventional Distributions

Friday
September 10th
11:50–12:15

Nikolaj Thams

University of Copenhagen

Abstract

A major question in causal inference is to predict what happens when we intervene on one or more of the variables, giving rise to an interventional distribution ([1]). Questions of interventional effects have gained significant interest due to constant appearances of interventions in real-world scenarios. In this work, we develop a framework for performing hypothesis tests in an interventional distribution, when we only have access to data from an (un-intervened) observational distribution. We resample the data with importance weights and use this to construct a test statistic for a hypothesis in the interventional distribution, which we show has the correct

asymptotic level. The need for hypothesis testing in interventional distributions is motivated by conditional independence tests, but we show that a range of other problems can be addressed by this framework.

Keywords: Hypthesis testing, Importance Resampling, Causality, Interventional distributions, Independence testing
AMS subject classifications: 62E99
Acknowledgements: This is joint work with Sorawit Saengkyongam, Niklas Pfister and Jonas Peters.

Bibliography

[1] Pearl, J. (2009). Causality. *Cambridge university press.*

Star-shaped order for distributions with multidimensional parameters and some applications

Idir Arab[1], Milto Hadjikyriakou[2], Paulo Eduardo Oliveira[1] and Beatriz Santos[1]

[1]*CMUC, Department of Mathematics, University of Coimbra, Portugal*
[2]*University of Central Lancashire, Cyprus*

Friday
September 10th
12:15–12:40

Abstract

The lifetime of complex systems with heterogeneous components are modelled by distributions depending on several parameters. We extend a criterion given by Saunders and Moran [4] allowing for the star-shaped comparison when multidimensional parameters are involved. This criterion is then applied to obtain star-shaped comparability within parallel and series systems for several underlying components behaviour, extending results proved by Kochar and Xu in [2, 3]. The ordering results we obtain contrast with the non-comparability results of the lifetime distributions of complex systems, proved by Arab et al. in [1], when considering ordering with respect to the stronger convex transform order.

Keywords: Star-shaped order, parallel systems, series systems.
AMS subject classifications: 60E15, 60E05, 62N05.
Acknowledgements: The authors IA, PEO, and BS are partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES. BS was also supported by

FCT, through the grant PD/BD/150459/2019, co-financed by the European Social Fund.

Bibliography

[1] Arab, I., Oliveira, P.E. and Hadjikyriakou M. (2020). Non comparability with respect to the convex transform order with applications. *J. Appl. Prob.* 57, 1339–1348.
[2] Kochar, S. and Xu, M. (2011). On the skewness of order statistics in multiple-outlier models. *J. Appl. Prob.* 48, 271–284.
[3] Kochar, S. and Xu, M. (2014).On the skewness of order statistics with applications. *Ann. Oper. Res.*, 212, 127–138.
[4] Saunders, I.W. and Moran, P.A.P. (1978). On the quantiles of the gamma and F distributions. *J. Appl. Prob.*, 15, 426–432.

An Application of a Geometric Process Model for Debugging and Testing Costs

Friday
September 10th
13:40–14:05

Mustafa Hilmi Pekalp[1] **and Halil Aydoğdu**[2]

[1]*Ankara University, Faculty of Applied Sciences, Department of Actuarial Sciences*
[2]*Ankara University, Faculty of Science, Department of Statistics*

Abstract

In this study, we propose a geometric process (GP) $\{N(t), t \geq 0\}$ to model the detection of software bugs where $N(t)$ denotes the number of the software bugs in $(0, t]$ for each fixed $t \geq 0$. Let the random variable (rv) $W_i, i = 1, 2, \ldots$ be the cost of fixing the ith software bug. Then, $W_i = c_0 + (i-1)U, i = 1, 2, \ldots$ where c_0 is a constant and U is a rv with mean c_1. The expected total debugging cost in $(0, t]$ is given by $C_1(t) = (\frac{2c_0 - c_1}{2})M_1(t) + \frac{c_1}{2}M_2(t), t \geq 0$ where $M_1(t)$ and $M_2(t)$ are the mean value and second moment functions of the GP, respectively. Also, let us assume that the cost of testing per unit time is a rv with mean c_2. Then, the total expected testing and debugging cost up to time t is $C_2(t) = tc_2 + C_1(t), t \geq 0$ [1].
To calculate the cost functions $C_1(t)$ and $C_2(t)$, it is necessary to compute the functions $M_1(t)$ and $M_2(t)$. In the literature, there are many papers on the calculation of these functions. However, the computation procedure depends on both distributional assumption of the first interarrival time of the GP and the estimations of the model and distribution parameters.
In this study, we consider the software system failure data given in [2] which is consistent with a GP model. For this data set, by applying the method given in [4], it can be shown that the data set can be modeled by a particular GP with a gamma distribution. Then, the model and distribution parameters can be estimated

by using the results of [3]. Numerical calculation proposed by [6] and [5] for the functions $M_1(t)$ and $M_2(t)$, respectively, can be used. Finally, we obtain the cost functions for the data set.

Keywords: Geometric process, geometric function, debugging and testing, software reliability
AMS subject classifications: 60K99, 62M99, 65Z05

Bibliography

[1] H. Pham, H. Wang. A Quasi-Renewal Process for Software Reliability and Testing Costs, IEEE Trans Syst Man Cybern.-Part A:Systems and Humans, 31:6:623-631, 2001.
[2] J.D. Musa, A. Iannino, K. Okumoto. Software Reliability: Measurement, Prediction, Application, McGraw-Hill, New York, 1987.
[3] J.S.K. Chan, Y. Lam, D.Y. Leung. Statistical Inference for Geometric Processes with Gamma Distributions. Computational Statistics and Data Analysis, 47-3, 565-581, 2004.
[4] M.H. Pekalp, H. Aydoğdu, K.F. Türkman. Discriminating Between Some Lifetime Distributions in Geometric Counting Processes. Communications in Statistics-Simulation and Computation, 10.1080/03610918.2019.1657452, 2019.
[5] M.H. Pekalp, H. Aydoğdu. An Integral Equation for the Second Moment Function of a Geometric Process and Its Numerical Solution. Naval Research Logistics, 65(2):176-184, 2018.
[6] Y. Tang, Y. Lam, Numerical Solution to an Integral Equation in Geometric Process, Journal of Statistical Computation and Simulation 77, 549-560, 2007.

Modelling block maxima with the blended generalised extreme value distribution

Silius M. Vandeskog[1], Sara Martino[1] and Daniela Castro-Camilo[2]

[1]*Norwegian University of Science and Technology*
[2]*University of Glasgow*

Friday
September 10th
14:05–14:30

Abstract

Short-term extreme precipitation can cause flash floods, large economic losses and immense destruction of infrastructure. In order to prepare for future extreme precipitation, we aim to apply a Bayesian hierarchical model for estimating return levels of the yearly maximum of sub-daily precipitation in Norway.
The generalised extreme value (GEV) distribution is a natural model for the yearly maximum of sub-daily precipitation. However, inference with the GEV distribution

is known to be difficult, partially because its support depends on its parameters. We propose to model the yearly maxima with the blended GEV (bGEV) distribution, which has the right tail of a Fréchet distribution and the left tail of a Gumbel distribution, resulting in a constant support that allows for more stable inference. Time series of sub-daily precipitation in Norway are short and noisy, making it difficult to place complex models on the parameters of the bGEV. We propose a new two-step model for block maxima that borrows strength from the peaks over threshold methodology by linking the scale parameter of a block maximum to the standard deviation of observations larger than some threshold. This increases the speed and stability of performing inference on the model. Inference speed is further increased by using the method of integrated nested Laplace approximations (INLA). Simulation studies are performed to test the two-step model. We find that the bGEV distribution allows for accurate predictions of large return levels and that the two-step model improves the performance when placing complex models on the parameters of the bGEV distribution with little available data.

Keywords: Extreme value theory, INLA, bGEV
AMS subject classifications: 60G70, 62F15, 62P12

The Limit Order Book statistical properties

Friday
September 10th
14:30–14:55

Dragana Radojičić

Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

Abstract

The main object of this research, namely the Limit Order Book (LOB), records all unexecuted buy and sell limit orders. We study the limit order book model described in terms of the mid-price process $S = \{S_t\}_{\{t\in\mathcal{T}\}}$ and by the two-parameter order volume process $\{V(t,p), t \in \mathcal{T}, p \in \mathbb{Z}\}$, where $\mathcal{T}$ is the set of all trade events during the selected day and $V(t,p)$ denotes the number of orders at time t awaiting execution at price level p. The order execution occurs when the mid-price reaches the price level at which there is some volume. Further, order cancellations are ignored, and it is assumed that no orders are placed at a distance L or larger than L from the mid-price. Thus, an order submitted at price level p is canceled at time t when $|S_t - p| \geq L$. Moreover, using the reflection principle (see Theorem 2 of III.7, P.89 in [1]) the distribution of the first passage time of the mid-price process through 1 is calculated under the assumption that the mid-price process can be approximated by the simple symmetric random walk.

Moreover, the probability generating function of the time to the next trade, and the probability generating function of the period of sequence of trade executions are derived. Further, the comparison of the distribution of the time to the next trade is observed under the two assumptions: orders are placed equidistantly in the LOB, and orders are placed in the LOB with respect to the geometric distribution.

Keywords: Limit Order Book, stochastic modeling, generating function, statistical analysis
AMS subject classifications: 60C05, 62P05, 62P20.

Bibliography

[1] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, New Jersey.

Mixed moment estimator for space-time heteroscedastic extremes: semi-parametric inference on extreme rainfall

Jessica Silva Lomba[1], **Maria Isabel Fraga Alves**[2] **and Cláudia Neves**[3]

[1,2]*CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal*
[3]*Department of Mathematics and Statistics, University of Reading, United Kingdom*

Friday
September 10th
14:55–15:20

Abstract

Extreme Value Theory provides the ideal framework for forecasting the frequency of extreme and hazardous events that are quite unlikely to occur and hard to predict. Within the general aim of extreme value statistics lies the estimation of probabilities of extreme events that have rarely been observed in the past, to which end the estimation of the so-called extreme value index is key. Due to accelerating climate change, extreme meteorological phenomena such as heavy precipitation, extreme temperature, strong winds and sea level rise, seem to be growing more severe and frequent, but the actual estimation of this evolution in extreme weather events remains subject to large uncertainty. Thus, inferential methods for the underlying non-stationary spatio-temporal processes are currently object of widespread interest. A recent development of extreme value theory is the concept of scedasis, through which a trend in extremes of time-space indexed observations can be captured and tactfully modelled within a semi-parametric framework.
In this talk we will look at how one can use series of data collected at several isolated observation points to model extremes of the whole space-time process in such a way

as to enable the mixed moment estimator of the extreme value index (introduced in [2]) to incorporate both space-time non-stationarity and dependence in a seamless way. Asymptotic properties (consistency and asymptotic normality) are worked out by building on the foundational work by Einmahl et al. [1] on the sequential tail empirical process and corresponding quantile process. Additionally, we look to expand this theory regarding empirical tail process in order to accommodate second order deterministic bias, a crucial term for estimating accuracy and performing bias reduction within the context of extreme value statistics. Finally, application of the extended mixed moment estimator is illustrated via long series of daily rainfall collected at gauging stations across the UK.

Keywords: Extreme value statistics, Extreme rainfall, Non-identical distributions, Semi-parametric inference, Sequential tail empirical process.
AMS subject classifications: 60F17, 60G15, 60G70, 62G05, 62G20, 62G30, 62G32, 62P12.
Acknowledgements: J. Silva Lomba and M.I. Fraga Alves' research is supported by Fundação para a Ciência e a Tecnologia, I.P. through PhD grant SFRH/BD/130764/2017 and project UIDB/00006/2020. C. Neves gratefully acknowledges support from EPSRC-UKRI Innovation Fellowship grant EP/S001263/1.

Bibliography

[1] Einmahl, J. H. J., Ferreira, A., de Haan, L., Neves, C. and Zhou, C. (2021). Spatial dependence and space-time trend in extreme events. *The Annals of Statistics*. To appear.
[2] Fraga Alves, M. I., Gomes, M. I., de Haan, L. and Neves, C. (2009). Mixed moment estimator and location invariant alternatives. *Extremes* 12, 149–185.

COVID-19: Study of the spread of the pandemic in Bulgaria

Friday
September 10th
17:00–17:25

Silvi-Maria Gurova

Institute for Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str., Block 25A, Sofia, Bulgaria

Abstract

Since the end of 2019, when the COVID-19 virus appeared in Wuhan, China and spread globally, the world has changed dramatically, as the pandemic affected many industries in each country, such as the economy, medicine, education and others. Scientists and medics around the world have started working to find the most appropriate drugs to prevent the infection, while governments are imposing restrictive

measures to limit the infection. Mathematicians focused on developing epidemiological mathematical models that describe the behavior of the infection and use them to predict the spread of COVID-19.
In this work, SEIRS (Susceptible-Exposed-Infectious-Recovered-Susceptible) epidemic model is presented which describes spread of the COVID-19 pandemic, not only among humans but also among animals [3]. Using real pandemic data for Bulgaria which we divide in two categories: Sofia city data and the Province data, we study if this model really describes pandemic process. We compare this model also with results using statistical packages from R such as ARIMA, Holt-Winters additive model (HWAAS) [1], [2].The results show that SEIRS model can be used successfully for study this pandemic disease and to predict behavior of the spread of epidemic in the province using the results for the spread of capital Sofia city.

Keywords: epidemic model, COVID-19, time-series, statistics, predictions.
AMS subject classifications: 62P10, 65C20
Acknowledgements: This work was supported by the project CoE on Informatics and ICT (grant No BG05M2OP001-1.001-0003) funded by the Operational Programme "Science and Education for Smart Growth" and also supported by the financial funds allocated to the Sofia University "St. Kl. Ohridski", grant No 80-10-87/2021.

Bibliography

[1] V. Papastefanopoulos, P. Linardatos, S. Kotsiantis, "Covid-19: A comparison of time series methods to forecast percentage of active cases per population", Appl. Sci. 2020, 10 (11), 3880; doi.org/10.3390/app10113880.
[2] S. Margenov, N. Popivanov, I. Ugrinova, St. Harizanov, and Tsv. Hristov, "Mathematical and Computer Modeling of COVID-19 Transmission Dynamics in Bulgaria by Time-depended Inverse SEIR Model", AIP Conference Proceedings 2333, 090024 (2021); https://doi.org/10.1063/5.0041868, https://arxiv.org/abs/2008.10360.
[3] S.-M. Gurova, "A Predator-Prey Model with SEIR and SEIRS Epidemic in the Prey", AMITANS'2019, AIP Conference Proceedings vol.2164, issue 1, pp. 080003-1-080015, 2019, DOI: 10.1063/1.5130826.

Modelling SARS-CoV2 transmission in a Bayesian framework

Petros Barmpounakis and Nikolaos Demiris

Athens University of Economics and Business

Friday
September 10th
17:25–17:50

Abstract

At the start of 2020 the emergence of Covid-19 pandemic brought unprecedented challenges and conditions across the globe. In the absence of pharmaceutical options, non-pharmaceutical interventions have been implemented in order to reduce the transmission. The reproduction number R_t, a way of quantifying transmissibility, has been a key part in statistical modelling and in assessing the efficiency of these interventions. However, due to severe under-reporting and bias of reported cases, the true prevalence of the disease remains unknown. We consider that the true infection dynamics can be described by a bayesian hierarchical non-linear epidemic model building on the work of Flaxman et al. We amend this model by inferring the location and magnitude of R_t changes, allowing the time when R_t changes to be stochastic and purely data driven, reducing the possibility of bias infliction. Our model draws information from the reported deaths which are likely less prone to under reporting to estimate the current infected cases through the infection fatality ratio. We study the cases of Greece and United Kingdom and estimate the instantaneous reproduction number R_t, as well as the cumulative number of infected individuals using Hamiltonian Monte Carlo in the Stan program. We compare our findings with antibody prevalence studies conducted in both countries to estimate the prevalence of Covid-19 like REACT2 study in England with the participation of 100000 adults, providing an external validation source of how the model performs and is capable of estimating the true number of infected individuals. We also further explore the potential influence of mobility data and rate of testing on the proportion of registered cases to the estimated ones using linear regression and bayesian neural networks.

Keywords: COVID-19, epidemic models, instantaneous reproduction number, Bayesian inference, Hamiltonian Monte Carlo
AMS subject classifications: 62F15

Bibliography

[1] Flaxman, S., Mishra, S., Gandy, A. *et al* (2020). *Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.*, Nature 584, 257–261.

[2] Helen Ward, Christina Atchison, Matthew Whitaker, Kylie EC Ainslie, Joshua Elliott, Lucy Okell, Rozlyn Redd, Deborah Ashby, Christl A Donnelly, Wendy Barclay, Ara Darzi, Graham Cooke, Steven Riley, Paul Elliott (2020). *Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults* medRxiv 2020.08.12.20173690.

A regime switching on Covid-19 analysis and prediction in Romania

Friday
September 10th
17:50–18:15

Marian Petrica[1], Radu D. Stochitoiu[2], Marius Leordeanu[3] and Ionel Popescu[4]

[1]*Faculty of Mathematics and Computer Science, University of Bucharest*

[2]*Faculty of Automatic Control and Computers, Polytechnic University of Bucharest*
[3]*Institute of Mathematics of the Romanian Academy and Polytechnic University of Bucharest*
[4]*Faculty of Mathematics and Computer Science, University of Bucharest and Institute of Mathematics of the Romanian Academy*

Abstract

Many problems in nature are modeled via a system of (partial) differential equations. For example the growth of a population, the spread of a virus, the evolution of weather, the heat distribution in a certain body, the flow of a fluid in a certain environment, models of dynamical systems, the evolution of the price of financial instruments and many other. Let's summarize this as a system of the form $X_t' = f(\alpha, X)$ where α is a set of parameters our model depends on. This equation can be also interpreted as a stochastic differential equation, but for the moment we will only consider the deterministic version of it.
In this framework there are two problems we want to treat. One is the determination of the parameters from a limited number of observations of the system at some given times. The other is the forecast on the system beyond the already observed data and eventually also prediction on what happened at intermediate times between the observations. This is a very basic problem
In this presentation we propose a regime separation for the analysis of Covid19 on Romania combined with mathematical models of SIR and SIRD. The main regimes we study are the free spread of the virus, the quarantine and partial relaxation and the last one is the relaxation regime. The main model we use is SIR which is a classical model, but because we can not fully trust the numbers of infected or recovered people we base our analysis on the number of deceased people which is more reliable. To actually deal with this we introduce a simple modification of the SIR model to account for the deceased separately. This in turn will be our base for fitting the parameters. We actually use the classical SIR model to detect the regime switching and in fact prove a proposition which shows that we can recover the parameters in a unique way from the daily observation of the number of infected and susceptible. This is the basis for guessing the main parameters in the model.
The actual estimation of the parameters in our SIRD model is done in two steps. The first one consists in training a neural network based on SIR models to detect the regime changes. Once this is done, we fit the main parameters of the SIRD model using a grid search near the values suggested by the neural network. At the end, we make some predictions on what the evolution will be in a timeframe of a month with the fitted parameters.

Keywords: statistical learning, Covid19, regimes, data
AMS subject classifications: 92B20, 62P35

Bibliography

[1] Janik Schüttler, Reinhard Schlickeiser, Frank Schlickeiser, and Martin Kröger, *Covid-19 predictions using a gauss model, based on data from april 2*, Physics **2** (2020), no. 2, 197–212.

[2] Calvin Tsay, Fernando Lejarza, Mark A Stadtherr, and Michael Baldea, *Modeling, state estimation, and optimal control for the us covid-19 outbreak*, arXiv preprint arXiv:2004.06291 (2020). wakefield2019spatio

[3] Jon Wakefield, Tracy Qi Dong, and Vladimir N Minin, *Spatio-temporal analysis of surveillance data*, Handbook of Infectious Disease Data Analysis (2019), 455–476.

Author index

Affiliation and Contacts

KEYNOTE SPEAKER	AFFILIATION	EMAIL ADDRESS
Christian H. Weiss	Helmut Schmidt University, Hamburg, Germany	weissc@hsu-hh.de
Ingrid Van Keilegom	KU Leuven, Leuven, Belgium	ingrid.vankeilegom@kuleuven.be
Markos Koutras	University of Piraeus, Piraeus, Greece	mkoutras@unipi.gr
Narayanaswamy Balakrishnan	McMaster University, Hamilton, Ontario, Canada	bala@mcmaster.ca
Sylvia Frühwirth-Schnatter	Vienna University of Economics and Business, Vienna, Austria	sylvia.fruehwirth-schnatter@wu.ac.at

COUNTRY	PARTICIPANT	AFFILIATION	EMAIL ADDRESS
Austria	Laura Vana	Vienna University of Economics and Business	laura.vana@wu.ac.at
	Peter Knaus	Vienna University of Economics and Business	peter.knaus@wu.ac.at
Belgium	Alexandre Jacquemain	UCLouvain	alexandre.jacquemain@uclouvain.be
	Negera Wakgari Deresa	KU Leuven	negerawakgari.deresa@kuleuven.be
Bulgaria	Silvi-Maria Gurova	IICT-BAS	smgurova@parallel.bas.bg
Croatia	Ivana Valentic	University of Zagreb	ivana.valentic@math.hr
	Zeljka Salinger	Cardiff University	salingerz@cardiff.ac.uk
Cyprus	Aimilia Savva	University of Cyprus	savva.emilia@ucy.ac.cy
	Eleni Manoli	University of Cyprus	manoli.eleni@ucy.ac.cy
Czech Republic	Jan Vávra	Charles University	vavraj@karlin.mff.cuni.cz
	Katerina Konasová	Charles University	konasova@karlin.mff.cuni.cz
Denmark	Nikolaj Thams	University of Copenhagen	nikolajthams@gmail.com
Finland	Janne Antero Kaseva	Natural Resources Institute Finland (LUKE)	janne.kaseva@luke.fi
	Joona Karjalainen	Aalto University	joona.karjalainen@aalto.fi
France	Mathilde Sautreuil	Université Paris-Saclay	mathilde.sautreuil@gmail.com
	Steven Golovkine	CREST, ENSAI	steven_golovkine@icloud.com
Georgia	Petre Babilua	Tbilisi State University	petre.babilua@tsu.ge
Germany	Henning Höllwarth	Technical University of Freiberg	henning.hoellwarth@math.tu-freiberg.de
	Laura Jula Vanegas	University of Goettingen	ljulava@gwdg.de

COUNTRY	PARTICIPANT	AFFILIATION	EMAIL ADDRESS
Greece	Emmanouil-Nektarios Kalligeris	University of the Aegean	ekalligeris@aegean.gr
	Petros Barmpounakis	Athens University of Economics and Business	barmpounakis@aueb.gr
Hungary	Máté Szalai	Bolyai Institute, University of Szeged	szalai.mate21@gmail.com
Israel	Royi Jacobovic	Hebrew university	royi.jacobovic@mail.huji.ac.il
	Yaniv Tenzer	Weizmann Institute of Science	yaniv.tenzer@gmail.com
Italy	Andrea Cappozzo	Politecnico di Milano	andrea.cappozzo@unimib.it
	Rosanna Cataldo	University of Naples Federico II	rosanna.cataldo2@unina.it
Lithuania	Jovita Gudan	Vilnius University	jovita.gudan@mif.vu.lt
	Saulius Jokubaitis	Vilnius University	saulius.jokubaitis@mif.vu.lt
Luxembourg	Yves Staudt	University of Applied Sciences of the Grisons	Yves.Staudt@fhgr.ch
Netherlands	Leo Huberts	University of Amsterdam	L.c.e.huberts@uva.nl
	Marina Tiana Dietrich	Vrije Universiteit Amsterdam	m.t.dietrich@vu.nl
Norway	Aliaksandr Hubin	University of Oslo	aliaksah@math.uio.no
	Silius Vandeskog	Norwegian University of Science and Technology	silius.m.vandeskog@ntnu.no
Poland	Krzysztof Rudaś	Polish Academy of Sciences	krzysztof.rudas@ipipan.waw.pl
	Mateusz Staniak	University of Wrocław	mateusz.staniak@uwr.edu.pl
Portugal	Beatriz Ferreira Santos	University of Coimbra	b14796@gmail.com
	Jessica Silva Lomba	Universidade de Lisboa	jslomba@fc.ul.pt
Romania	Luigi-Ionut Catana	University of Bucharest	luigi_catana@yahoo.com
	Marian Petrica	University of Bucharest	marianpetrica11@gmail.com
Russia	Mariia Platonova	Saint–Petersburg State University	mariyaplat@gmail.com
	Nikolai Alekseevich Slepov	Lomonosov Moscow State University	naslepov@mail.ru
Serbia	Dragana Radojicic	Mathematical Institute of the Serbian Academy of Sciences and Arts	gagaradojicic@gmail.com
	Zoran Cedo Vidovic	University of Belgrade	zoranvidovic1990@gmail.com
Slovakia	Matúš Padyšák	Comenius University	matus.padysak@fmph.uniba.sk
Slovenia	Andrej Srakar	University of Ljubljana	andrej.srakar@ier.si
	Jakob Peterlin	University of Ljubljana	jakob.peterlin@mf.uni-lj.si
Spain	Elena Castilla González	Complutense University of Madrid	elecasti@ucm.es
	Roi Naveiro	Institute of Mathematical Sciences (ICMAT-CSIC)	roi.naveiro@icmat.es
Sweden	Johan Larsson	Lund University	johan.larsson@stat.lu.se
Switzerland	Martin Bladt	University of Lausanne,	martinbladt@gmail.com
	Simone Tiberi	University of Zurich	simone.tiberi@uzh.ch
Turkey	Murat Ozkut	Izmir University of Economics	murat.ozkut@ieu.edu.tr
	Mustafa Hilmi Pekalp	Ankara University	mpekalp@ankara.edu.tr
UK, England	Chengchun Shi	London School of Economics and Political Science	c.shi7@lse.ac.uk
	Marina Riabiz	King's College London	marina.riabiz@gmail.com

COUNTRY	PARTICIPANT	AFFILIATION	EMAIL ADDRESS
UK, Northern Ireland	Laura Boyle	Queen's University Belfast	laura.boyle@qub.ac.uk
	Nicola-Ann Jean Stevens	Queen's University Belfast	nstevens01@qub.ac.uk

Sponsors

Πάντειον Πανεπιστήμιο Κοινωνικών & Πολιτικών Επιστημών
Panteion University of Social and Political Sciences

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY
ΜΕΤΑΠΤΥΧΙΑΚΟ ΣΤΑΤΙΣΤΙΚΗ
MSc IN STATISTICS

ΕΚΔΟΣΕΙΣ GUTENBERG

Publications PROPOBOS

Auspices

RESURGO MUTATA EADEM
Bernoulli Society for Mathematical Statistics and Probability