18th European Young Statisticians Meeting

26-30 August 2013, Osijek, Croatia

Book of Abstracts

Preface

This booklet contains basic information about the *18th European Young Statisticians Meeting* (18th EYSM) to be held at the Department of Mathematics, J.J. Strossmayer University of Osijek, from 26 to 30 August 2013.

There are twenty five European countries participating at the 18th EYSM. The representative of each country in the International Organizing Committee (IOC) was responsible for inviting at most two young scientists from their country whose research interests are in the stochastic range, from pure probability theory to applied statistics. In this text the term ”young scientist” refers to scientists who are younger than thirty or have up 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 are giving 60-minute keynote lectures
- forty nine invited young scientists are having 20-minute talks.

Traditional for European Young Statisticians Meetings, there are no parallel sections. However, the talks of invited young scientists are divided into the following sessions:

1. statistical inference
    - estimation
    - testing procedures
2. statistical applications
    - biology and medicine
    - economics and insurance
    - image analysis
    - engineering, industry and seismology
3. theory of continuous time stochastic processes
4. inequalities and stochastic orderings

5. diagnostics and decision theory
6. optimal design
7. other topics in statistics and probability.

All the participants have an opportunity to publish a short paper, i.e. extended abstract of their lectures, in the Proceedings of the 18th EYSM. The review process for short papers is organized by the IOC. The IOC representative of each country proposes reviewers for papers of participants from their country or review the papers personally.

This booklet, beside all important information about the Conference, contains the scientific program, abstracts of all lectures and list of participants together with their affiliations and contacts. Abstracts of all contributors follow the schedule of lectures from the scientific program.

More details about the *18th European Young Statisticians Meeting* could be found at the website www.mathos.unios.hr/eysm18.

18th European Young Statisticians Meeting

Organizer

Department of Mathematics, J.J. Strossmayer University of Osijek

Auspices

Bernoulli Society for Mathematical Statistics and Probability

J.J. Strossmayer University of Osijek

International Organizing Committee

Jürgen Pilz, Alpen-Adria Universitaet Klagenfurt, Austria

Gregor Kastner, Vienna University of Economics and Business, Austria

Johan Segers, Université catholique de Louvain, Belgium

Metodi Nikolov, University of Sofia, Bulgaria

Marek Dvořák, Charles University in Prague, Czech Republic

Mirta Benšić, J.J. Strossmayer University of Osijek, Croatia

Michael Sørensen, University of Copenhagen, Denmark

Paavo Salminen, Abo Akademi University, Finland

Olivier Bouaziz, Universite Paris Descartes Paris 5, France

Andrea Krajina, Georgia Augusta University Göttingen, Germany

Ioannis Ntzoufras, Athens University of Economics and Business, Greece

Gyorgy Terdik, University of Debrecen, Hungary

Emanuele Taufer, University of Trento, Italy

Francesca Ieva, Politecnico di Milano, Italy

Harry van Zanten, University of Amsterdam, The Netherlands

Grzegorz Wyłupek, University of Wrocaw, Poland

Paulo C. Rodrigues, Nova University of Lisbon, Portugal

Eugenia Panaitescu, "Carol Davila" Univ. of Medicine and Pharmacy, Romania

Alexey Muravlev, Steklov Mathematical Institute, Russia

Vladimír Lacko, Comenius University, Slovakia

Aleš Toman, University of Ljubljana, Slovenia

Alba Maria Franco Pereira, University of Vigo, Spain

Silvelyn Zwanzig, Uppsala University, Sweden

Johanna Ziegel, University of Bern, Switzerland

Deniz Inan, Marmara University, Turkey

Ludmila Sakhno, Taras Shevchenko National University of Kyiv, Ukraine

Steven Gilmour, University of Southampton, United Kingdom

Nikolai N. Leonenko, Cardiff University, United Kingdom

Local Organizing Committee

Mirta Benšić, J.J. Strossmayer University of Osijek

Danijel Grahovac, J.J. Strossmayer University of Osijek

Petra Posedel, Zagreb School of Economics and Management

Nenad Šuvak, J.J. Strossmayer University of Osijek

Keynote speakers

Bojan Basrak, University of Zagreb, Croatia

Nikolai N. Leonenko, Cardiff University, United Kingdom

Jürgen Pilz, Alpen-Adria Universitaet Klagenfurt, Austria

Johan Segers, Université catholique de Louvain, Belgium

Michael Sørensen, University of Copenhagen, Denmark

Conference structure: keynote lectures, invited lectures

Conference language: English

Scientific Program

Monday – August 26, 2013

9:00 - 9:30	**Registration**
9:30 - 10:00	**Opening ceremony**
10:00 - 11:00	**Keynote lecture**

Nikolai N. Leonenko
School of Mathematics, Cardiff University, United Kingdom
Multifractal products of geometric stationary processes

11:00 - 11:30 **Coffee break**

11:30 - 13:10 **Session: Statistical inference - estimation**
Chairman: Nikolai N. Leonenko

11:30 – 11:55	Kostiantyn Ralchenko *Drift parameter estimation in models with fractional Brownian motion by discrete observations*
11:55 – 12:20	Lauri Viitasaari *Parameter estimation for discretely observed fractional Ornstein-Uhlenbeck process of the second kind*
12:20 – 12:45	Danijel Grahovac *On the tail index inference based on scaling functions method*
12:45 – 13:10	Juan Juan Cai *Estimation of extreme risk regions under multivariate regular variation*

13:10 – 15:00 **Lunch**

15:00 - 16:40	**Session: Statistical inference - testing procedures** **Chairman:** Emanuele Taufer	
	15:00 – 15:25	Philip Preuß *Discriminating between long-range dependence and non-stationarity*
	15:25 – 15:50	Lukas Steinberger *Statistical inference when fitting simple models to high dimensional data*
	15:50 – 16:15	Katarína Starinská *Score test statistic for change-point detection in AR time series with dependent errors*
	16:15 – 16:40	Kinga Sikolya *Stability testing for a special Pickard model*
16:40 – 17:00	**Coffee break**	
17:00 - 18:40	**Session: Statistical applications - biology and medicine** **Chairman:** Bojan Basrak	
	17:00 – 17:25	Francesco Stingo *An integrative Bayesian modeling approach to imaging genetics*
	17:25 – 17:50	Piotr Szulc *Model selection approach for genome wide association studies in admixed populations*
	17:50 – 18:15	Mircea Dumitru *Modeling the link between two genes expressions and the toxicity of some chemotherapy drugs for chronotherapy of cancer*
	18:15 – 18:40	Alicja Szabelska *Statistical methods for RNAseq coverage profiles and their applications in biological phenomena*

Tuesday – August 27, 2013

9:00 - 9:50 **Session: Statistical applications - economics and insurance**
Chairman: Petra Posedel

9:00 – 9:25 Žiga Kotnik
Statistical methodology in the scope of performance budgeting

9:25 – 9:50 Gábor Szűcs
Stochastic interest rates in life insurance mathematics

9:50 - 10:00 **Break**

10:00 - 11:00 **Keynote lecture**

Michael Sørensen
Department of Mathematical Sciences
University of Copenhagen, Denmark
Statistics for stochastic differential equations - Two approaches

11:00 - 11:30 **Coffee break**

11:30 - 13:10 **Session V: Statistical inference - estimation**
Chairman: Michael Sørensen

11:30 – 11:55 Snježana Lubura
Analysis of the approximate maximum likelihood estimators of diffusion parameters by simulations

11:55 – 12:20 Alexey Doronin
Adaptive estimation in mixture models with varying mixing probabilities

12:20 – 12:45 Manuela Cattelan
Models for dependent paired comparison data

12:45 – 13:10 Rudolf Schenk
Adaptive nonparametric Bayesian estimation for Gaussian inverse regression

13:10 – 15:00 **Lunch**

15:00 - 16:40	**Session: Inequalities and stochastic ordering** **Chairman:** Nikolai N. Leonenko	
	15:00 – 15:25	Abakirova Aygul *Stochastic versions of logarithmic Sobolev inequalities*
	15:25 – 15:50	Ekaterina Krymova *On oracle inequality for exponential weighting of ordered smoothers*
	15:50 – 16:15	Nuria Torrado-Robles *Upper and lower bounds for ordered random variables*
	16:15 – 16:40	Mircea Dragulin *Mixture Lorenz curves. Three new models*
16:40 – 17:00	**Coffee break**	
17:00	**Guided city tour**	

Wednesday – August 28, 2013

9:00 - 9:50	**Session: Statistical applications - image analysis** **Chairman:** Mirta Benšić
	9:00 – 9:25 Leena Annukka Pasanen *Bayesian multiscale analysis of images*
	9:25 – 9:50 Çigdem Güngör *The comparison of Harris corner detection method and statistical methods for detection of dominant points of two dimensional objects*
9:50 - 10:00	**Break**
10:00 - 11:00	**Keynote lecture**
	Jürgen Pilz Institute of Statistics, University of Klagenfurt, Austria *Some advances in Bayesian spatial prediction and sampling design*
11:00 - 11:30	**Coffee break**
11:30 - 13:10	**Session: Statistical inference - estimation** **Chairman:** Jürgen Pilz
	11:30 – 11:55 Eric Sibony *Multiresolution method for ranking*
	11:55 – 12:20 Sarah Ouadah *Uniform-in-bandwidth kernel estimation for censored data*
	12:20 – 12:45 Petro Kolesnyk *Bi-log-concave distribution functions and confidence bands*
	12:45 – 13:10 Botond Szabo *Adaptive confidence sets from Bayes procedures*
13:10 – 15:00	**Lunch**

15:00 - 16:40	**Session: Statistical applications - biology and medicine** **Chairman:** Miljenko Huzak	
	15:00 – 15:25	Panagiotis Papastamoulis *Approximating the posterior distribution of mixture weights with application to transcript expression estimation*
	15:25 – 15:50	Joke Durnez *Alternative based thresholding for pre-surgical fMRI*
	15:50 – 16:15	Ferenc Rárosi *Regression based predictions for irradiation doses*
	16:15 – 16:40	Laetitia Da Costa Teixeira *Competing risks analysis in nephrology research: an example in peritoneal dialysis*
16:40 – 17:00	**Coffee break**	
17:00 - 18:40	**Session: Statistical applications - engineering, industry and seismology** **Chairman:** Nenad Šuvak	
	17:00 – 17:25	Petr Novak *Accelerated failure time model for repairable systems*
	17:25 – 17:50	Peter Scheibelhofer *Robust multivariate process control of multiway data with root cause analysis*
	17:50 – 18:15	Guillermo Vinué *Developing statistical methodologies for anthropometry*
	18:15 – 18:40	Katerina Orfanogiannaki *Hidden Markov models in modeling time series of earthquakes*

Thursday – August 29, 2013

9:00 - 9:50 **Session: Diagnostics and decision theory**
Chairman: Bojan Basrak

9:00 – 9:25 Emmanuel Ogundimu
On sample selection models and skew distributions

9:25 – 9:50 Ufuk Beyestaş
BCa JaB method as a diagnostic tool for linear regression models

9:50 - 10:00 **Break**

10:00 - 11:00 **Keynote lecture**

Johan Segers
Institut de statistique, biostatistique et sciences actuarielles, Universit catholique de Louvain, Belgium
Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

11:00 - 11:30 **Coffee break**

11:30 - 13:10 **Session: Theory of continuous time stochastic processes**
Chairman: Johan Segers

11:30 – 11:55 Maik Görgens
The zero area Brownian bridge

11:55 – 12:20 Alexander Sokol
Intervention in Ornstein-Uhlenbeck SDEs

12:20 – 12:45 Andreas Basse-O'Connor
Some properties of a class of continuous time moving average processes

12:45 – 13:10 Kaspar Stucki
Gibbs point process approximation based on Steins method

13:10 **Lunch, excursion and conference dinner**

Friday – August 30, 2013

9:00 - 9:50 **Session: Optimal design**
Chairman: Petra Posedel

9:00 – 9:25 Verity Fisher
Optimal designs for discriminating between functional linear models

9:25 – 9:50 Alena Bachratá
A stochastic optimization method for constructing optimal block designs with linear constraints

9:50 - 10:00 **Break**

10:00 - 11:00 **Keynote lecture**

Bojan Basrak
Department of Mathematics, University of Zagreb, Croatia
On dependent regularly varying observations

11:00 - 11:30 **Coffee Break**

11:30 - 13:10 **Session: Other topics in statistics and probability**
Chairman: Nenad Šuvak

11:30 – 11:55 Katja Trinajstić
Directed random graphs and convergence to the Tracy-Widom distribution

11:55 – 12:20 B.G. Manjunath
Some remarks on normal conditionals and normal projections

12:20 – 12:45 Eike Christian Brechmann
Constructing hierarchical copulas using the Kendall distribution function

12:45 – 13:10 Bono Nonchev
Minimum description length principle and distribution complexity of spherical distributions

13:10 – 13:35 Teodosi Geninski
Skewed sub-Gaussian multivariate distribution

13:35 – 14:00 **Closing ceremony**

14:00 **Lunch**

Contents

18th European Young Statisticians Meeting

Keynote lectures

Multifractal products of geometric stationary processes

Nikolai N. Leonenko

School of Mathematics, Cardiff University, United Kingdom

Abstract

This is joint work with D. Denisov (Cardiff University).
Multifractal and monofractal models have been used in many applications in hydrodynamic turbulence, finance, genomics, computer network traffic, etc. (see, for example, [7]). There are many ways to construct random multifractal models ranging from simple binomial cascades to measures generated by branching processes and the compound Poisson process ([4] - [7]).
Anh, Leonenko and Shieh ([1]-[3]) and Leonenko and Shieh [8] considered multifractal products of stochastic processes as defined in [**?**], but they provide a new interpretation of the conditions on the characteristics of geometric stationary processes in terms of the moment generating functions.
We investigate the properties of multifractal products of geometric Gaussian processes with possible long-range dependence and geometric Ornsteinf-Uhlenbeck processes driven by Lévy motion and their finite and infinite superpositions. We present the general conditions for the $\mathcal{L}_q$ convergence of cumulative processes to the limiting processes and investigate their q-th order moments and Rényi functions, which are nonlinear, hence displaying the multifractality of the processes as constructed. We also establish the corresponding scenarios for the limiting processes, such as log-normal, log-gamma, log-tempered stable or log-normal tempered stable scenarios.

Bibliography

[1] Anh, V. V., Leonenko, N. N. and Shieh, N.-R. (2008). Multifractality of products of geometric Ornstein-Uhlenbeck-type processes. *Adv. in Appl. Probab.* **40** 1129–1156.

[2] Anh, V. V., Leonenko, N. N. and Shieh, N.-R. (2009). Multifractal scaling of products of birth-death processes. *Bernoulli* **15** 508–531.

[3] Anh, V. V., Leonenko, N. N., Shieh, N.-R. and Taufer, E. (2010). Simulation of multifractal products of Ornstein-Uhlenbeck type processes. *Nonlinearity* **23** 823–843.

[4] Bacry, E. and Muzy, J.F. (2003). Log-infinitely divisible multifractal processes. *Comm. Math. Phys.* 236 (2003), 449–475.

[5] Barndorf-Nilsen, O.E. and Shmigel, Yu (2004). Spatio-temporal modeling based on Lévy processes, and its applications to turbulence. (Russian) *Uspekhi Mat. Nauk* 59, 63–90; translation in *Russian Math. Surveys* 59, 65–90.

[6] Denisov, D. and Leonenko, N. (2011). Multifractality of products of geometric stationary processes. *Submitted*, published in arxiv.org/abs/1110.2428.

[7] Doukhan, P., Oppenheim, G. and Taqqu, M.S.(2003). *Theory and Applications of Long-range Dependence.* Birkhäuser Boston.

[8] Leonenko, N.N and Shieh N.-R. (2013). Rényi function for multifractal random fields. *Fractals*, in press.

Statistics for stochastic differential equations - Two approaches

Michael Sørensen

Department of Mathematical Sciences, University of Copenhagen, Denmark

Abstract

For discrete-time observations of the solution to a stochastic differential equation, there is usually no explicit expression for the likelihood function, which is a product of transition densities. Therefore, the likelihood function must be approximated. A brief review will be given of a broad spectrum of approximation methods. Two approaches will be presented in detail. Martingale estimating functions are a simple way of approximating likelihood inference that provides estimators which are easy to calculate. These estimators are generally consistent, and if the estimating function is chosen optimally, they are efficient in a high frequency asymptotic scenario, where the sampling frequency goes to infinity. At low sampling frequencies, efficient estimators can be obtained by more accurate approximations to likelihood inference based on simulation methods, including both the stochastic EM-algorithm and Bayesian approaches like the Gibbs sampler. These methods are much more computer intensive. Simulation of diffusion bridges plays a central role. Therefore this highly non-trivial problem has been investigated actively over the last 10 years. A simple method for diffusion bridge simulation will be presented and applied to likelihood inference for stochastic differential equations.

Some advances in Bayesian spatial prediction and sampling design

Jürgen Pilz

Alpen-Adria Universitaet Klagenfurt, Austria

Abstract

In my talk, I will report on recent work with my colleagues G. Spoeck and H. Kazianka in the area of Bayesian spatial prediction and design [1]-[4].

The Bayesian approach not only offers more flexibility in modeling but also allows us to deal with uncertain distribution parameters, and it leads to more realistic estimates for the predicted variances. We report on some experiences gained with our approach during a European project on "Automatic mapping of radioactivity in case of emergency".
We then go on and apply copula methodology to Bayesian spatial modeling and derive predictive distributions. Moreover, I report on recent results on finding objective priors for the crucial nugget and range parameters of the widely used Matern-family of covariance functions.
Further on, I briefly consider the challenges in stepping from the purely spatial setting to spatio-temporal modeling and prediction.

Finally, I will consider the problem of choosing an "optimal" spatial design, i.e. finding an optimal spatial configuration of the observation sites minimizing the total mean squared error of prediction over an area of interest. Using Bessel-sine/cosine-expansions for random fields we arrive at a design problem which is equivalent to finding optimal Bayes designs for linear regression models with uncorrelated errors, for which powerful methods and algorithms from convex optimization theory are available. I will also indicate problems and challenges with optimal Bayesian design when dealing with more complex design criteria such as minimizing the averaged expected lengths of predictive intervals over the area of interest.

Bibliography

[1] H. Kazianka and J. Pilz (2011). Bayesian spatial modeling and interpolation using copulas. *Computers & Geosciences.* 37(3): 310-319.
[2] H. Kazianka and J. Pilz (2012). Objective Bayesian analysis of spatial data taking account of nugget and range parameters. *The Canadian Journal of Statistics.* 40(2): 304-327.
[3] J. Pilz, H. Kazianka and G. Spoeck (2012). Some advances in Bayesian spatial prediction and sampling design. *Spatial Statistics.* 1: 65-81.
[4] G. Spoeck and J. Pilz (2013). Spatial sampling design based on spectral approximations of the error process. In: *Spatio-temporal design: Advances in Efficient Data Acquisition* (W.G. Mueller and J. Mateu, Eds.), Wiley, New York, 72-102

Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation

Johan Segers[1], **Ramon van den Akker**[2] **and Bas Werker**[2]

[1] *Université catholique de Louvain, Belgium*
[2] *Tilburg University, The Netherlands*

Abstract

For multivariate Gaussian copula models with unknown margins and general correlation structures, a simple, rank-based and semiparametrically efficient estimator is proposed. An algebraic representation of relevant subspaces of the tangent space is constructed that allows to easily study questions of adaptivity with respect to the unknown marginal distributions and of efficiency of the pseudo-likelihood estimator and the normal-scores rank correlation coefficient. Some well-known examples are treated explicitly: circular correlation matrices, factor models, and Toeplitz matrices, special cases being exchangeable structures, moving average models and autoregressive models. For constructed examples, the asymptotic relative efficiency of the pseudo-likelihood estimator can be as low as 20 percent. For finite samples, these findings are confirmed by Monte Carlo simulations.

On dependent regularly varying observations

Bojan Basrak

Department of Mathematics, University of Zagreb, Croatia

Abstract

It is well known that the extremal behavior of stationary sequences can be nicely captured using the language of point processes. We explain how this theory extends from iid to dependent sequences as long as this dependence disappears in time. The theory turns out to be especially elegant when applied to stationary regularly varying sequences, which we discuss in detail.

In particular, the dependence structure of extremes for such sequences can be described using the concept of the tail process. By application of the point processes theory, this leads to various asymptotic results for extremes and sums of such sequences, including some nonstandard functional limit theorems.

18th European Young Statisticians Meeting

Abstracts

Drift parameter estimation in models with fractional Brownian motion by discrete observations

Kostiantyn Ralchenko[1]

[1] *Taras Shevchenko National University of Kyiv*

Abstract

Let $B^H = \{B^H_t, t \geq 0\}$ be fractional Brownian motion with Hurst index $H \in (1/2, 1)$. We study the problem of estimating of unknown drift parameter from [1]. Consider the stochastic differential equation driven by fractional Brownian motion B^H:

$$\begin{aligned} dX_t &= \theta a(X_t)dt + b(X_t)dB^H_t, \quad 0 \leq t \leq T, \quad T > 0, \\ X\big|_{t=0} &= X_0 \in \mathbb{R}. \end{aligned} \tag{1}$$

Here $\theta \in \mathbb{R}$ is unknown parameter to be estimated.
Suppose that we observe values $X_{\frac{k}{2^n}}$, $k = 0, 1, \ldots, 2^{2n}$. Consider two estimators for θ:

$$\hat{\theta}^{(1)}_n = \frac{\sum_{k=1}^{2^{2n}} \left(\frac{k}{2^n}\right)^{-\alpha} \left(2^n - \frac{k}{2^n}\right)^{-\alpha} b^{-1}\left(X_{\frac{k-1}{2^n}}\right)\left(X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\right)}{\sum_{k=1}^{2^{2n}} \left(\frac{k}{2^n}\right)^{-\alpha} \left(2^n - \frac{k}{2^n}\right)^{-\alpha} b^{-1}\left(X_{\frac{k-1}{2^n}}\right) a\left(X_{\frac{k-1}{2^n}}\right)\frac{1}{2^n}},$$

$$\hat{\theta}^{(2)}_n = \frac{\sum_{k=1}^{2^{2n}} b^{-1}\left(X_{\frac{k-1}{2^n}}\right)\left(X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\right)}{\sum_{k=1}^{2^{2n}} b^{-1}\left(X_{\frac{k-1}{2^n}}\right) a\left(X_{\frac{k-1}{2^n}}\right)\frac{1}{2^n}}.$$

In the simplest cases (for example, $a = b$) the estimator $\hat{\theta}^{(1)}_n$ coincides with a discrete version of maximum-likelihood estimator. $\hat{\theta}^{(2)}_n$ is a non-standard estimator. We prove that both estimators converge to the true value of the parameter θ.

Keywords: fractional Brownian motion, stochastic differential equation, parameter estimation, strong consistency, discretization.
AMS subject classifications: Primary: 60G22, 62F10. Secondary: 60H10, 62F12.

Bibliography

[1] Mishura, Y. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*, Lecture Notes in Mathematics, vol. 1929, Springer, Berlin.

Parameter estimation for discretely observed fractional Ornstein-Uhlenbeck process of the second kind

Lauri Viitasaari[1] **and Ehsan Azmoodeh**[2]

[1] *Department of Mathematics and System Analysis, Aalto University School of Science, Finland*
[2] *Faculté des Sciences, de la Technologie et de la Communication, Université du Luxembourg, Luxembourg*

Abstract

Let W_t be a standard Brownian motion. Classical Ornstein-Uhlenbeck processes can be obtained via solution of the Langevin equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}W_t$ or via Lamperti transform $X_t^{(\alpha)} = e^{-\theta t} W_{\alpha e^{2\theta t}}$. With a particular choice $\alpha = \frac{1}{2\theta}$ these two processes are the same in a sense that they have the same finite dimensional distributions. This is not the case however if one replaces Brownian motion W_t with fractional Brownian motion B_t^H. In particular, the process arising from Lamperti transform can be viewed as a solution to Langevin type equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}Y_t$ with a noise Y given by

$$Y_t = \int_0^t e^{-s} \mathrm{d}B^H_{He^{\frac{t}{H}}}.$$

As a result we obtain two different fractional Ornstein-Uhlenbeck processes depending on the approach. The solution to Langevin equation $\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \mathrm{d}B_t^H$ is referred to fractional Ornstein-Uhlenbeck process of the first kind and the process arising from Lamperti transform is referred to fractional Ornstein-Uhlenbeck process of the second kind.

An interesting problem in mathematical statistics is to estimate the unknown parameter θ. One approach is to consider LSE estimator based on Skorokhod integrals. This is considered by Hu and Nualart [3] for first kind process and by Azmoodeh and Morlanes [1] for second kind process. However, divergence integrals cannot be computed from the path of the process. Another approach is to observe the path of the process and estimate the unknown parameter θ directly from the observations. We consider discretely observed second kind process and find strongly consistent estimator. We also introduce an estimator based on generalised quadratic variations for Hurst parameter H. Moreover, we derive central limit theorems for our estimators. Similar results for first kind process is derived by Brouste and Iacus [2].

Keywords: fractional Ornstein-Uhlenbeck processes, Langevin equation, parameter estimation, central limit theorem
AMS subject classifications: 60G22, 60H07, 62F10, 62F12
Acknowledgements: Lauri Viitasaari thanks the Finnish Doctoral Programme in Stochastics and Statistics for financial support.

Bibliography

[1] Azmoodeh, E. and Morlanes, I. (2012). Drift parameter estimation for Fractional Ornstein-Uhlenbeck process of the second kind. *Manuscript.*

[2] Brouste, A. and Iacus, S. M. (2012). Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Computational Statistics.* DOI 10.1007/s00180-012-0365-6.

[3] Hu, Y. and Nualart, D. (2010). Parameter estimation for fractional Ornstein-Uhlenbeck processes. *Statist. Probab. Lett.* 80, no. 11-12, 1030-1038.

On the tail index inference based on scaling functions method

Danijel Grahovac[1], Mofei Jia[2], Nikolai Leonenko[3] and Emanuele Taufer[2]

[1]*Department of Mathematics, Josip Juraj Strossmayer University, Osijek, Croatia*
[2]*Department of Economics and Management, University of Trento, Italy*
[3]*Cardiff School of Mathematics, Cardiff University, UK*

Abstract

In [3] a new method has been presented for making inference about the tail of unknown heavy-tailed distribution. Method is is based on the asymptotic properties of the empirical structure function, a variant of statistic that resembles usual sample moments. Using this approach one can successfully inspect the nature of the tail of the underlying distribution, as well as provide estimates on the unknown tail index. Here we briefly describe the method and test its performance on some simulated and real world data by comparing it with the well known Hill estimator.

Keywords: heavy-tailed distributions, tail index, empirical structure function, scaling functions, Hill estimator.
AMS subject classifications: 62F10, 62F12, 62E20.

Bibliography

[1] De Haan, L., & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

[2] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*, vol. 33. Springer Verlag.

[3] Grahovac, D., Jia, M., Leonenko, N., & Taufer, E. (2013). Tail index estimation based on the asymptotic properties of the empirical structure function. *Submitted.*

[4] Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394–419.
[5] Meerschaert, M., & Scheffler, H. (1998). A simple robust estimation method for the thickness of heavy tails. *Journal of Statistical Planning and Inference*, *71*(1), 19–34.

Estimation of extreme risk regions under multivariate regular variation

Juan-Juan Cai[1] **and John H.J. Einmahl**[1] **and Laurens de Haan**[2]

[1] *Tilburg University*
[2] *Tilburg University, Lisbon University, Erasmus University, Rotterdam*

Abstract

When considering d possibly dependent random variables one is often interested in extreme risk regions, with very small probability p. We consider risk regions of the form $\{\mathbf{z} \in \mathbb{R}^d : f(\mathbf{z}) \leq \beta\}$, where f is the joint density and β a small number. Estimation of such an extreme risk region is difficult since it contains hardly any or no data. Using extreme value theory, we construct a natural estimator of an extreme risk region and prove a refined form of consistency, given a random sample of multivariate regularly varying random vectors. In a detailed simulation and comparison study the good performance of the procedure is demonstrated. We also apply our estimator to financial data.

Keywords: Extremes, level set, multivariate quantile, rare event, spectral density
AMS subject classifications: 62G32, 62G05, 62G079.

Discriminating between long-range dependence and non-stationarity

Philip Preuß[1] and Mathias Vetter[1]

[1] *Ruhr-University Bochum*

Abstract

This paper is devoted to the discrimination between a stationary long-range dependent model and a non stationary process. We develop a nonparametric test for stationarity in the framework of locally stationary long memory processes which is based on a Kolmogorov-Smirnov type distance between the time varying spectral density and its best approximation through a stationary spectral density. We show that the test statistic converges to the same limit as in the short memory case if the (possibly time varying) long memory parameter is smaller than 1/4 and justify why the limiting distribution is different if the long memory parameter exceeds this boundary. Concerning the latter case the novel FARI(∞) bootstrap is introduced which provides a bootstrap-based test for stationarity that only requires the long memory parameter to be smaller than 1/2 which is the usual restriction in the framework of long-range dependent time series. We investigate the finite sample properties of our approach in a comprehensive simulation study and apply the new test to a data set containing log returns of the S&P 500.

Keywords: Empirical spectral measure, Integrated periodogram, Locally stationary process, Long Memory, Spectral density
AMS subject classifications: 62M10
Acknowledgements: This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG).

Statistical inference when fitting simple models to high dimensional data

Lukas Steinberger[1] **and Hannes Leeb**[1]

[1] *University of Vienna, Department of Statistics and OR*

Abstract

We study linear subset regression in the context of the high-dimensional overall model $y = \theta' Z + u$ with univariate response y and a $d-$vector of random regressors Z, independent of u. Here, 'high-dimensional' means that the number n of available observations may be much less than d. We consider simple linear submodels where y is regressed on a set of p regressors given by $X = B'Z$, for some $d \times p$ matrix B with $p \leq n$. The corresponding simple model, i.e., $y = \gamma' X + v$, can be justified by imposing appropriate restrictions on the unknown parameter θ in the overall model; otherwise, this simple model can be grossly mis-specified. We show that the least-squares predictor obtained by fitting the simple linear model is typically close to the Bayes predictor $E[y|X]$ in a certain sense, uniformly in $\theta \in \mathbb{R}^d$, provided only that d is large. Moreover, we establish the asymptotic validity of the standard $F-$test on the surrogate parameter which realizes the best linear population level fit of X on y, in an appropriate sense. On a technical level, we extend recent results from [4] on conditional moments of projections from high-dimensional random vectors; see also [1, 2, 3].

Keywords: High-dimensional models, mis-specified model, regression analysis, prediction, F-test.
AMS subject classifications: 62F05, 62J05

Bibliography

[1] Diaconis, P. and Freedman, D. (1984). Asymptotics of Graphical Projection Pursuit. *The Annals of Statistics* **12**, 3, 793–815.

[2] Dümbgen, L. and Conte-Zerial, P. (2012). On Low-Dimensional Projections of High-Dimensional Distributions. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. A Festschrift in Honor of Jon Wellner*, M. Banerjee; F. Bunea; J. Huang; V. Koltchinskii; and M.H. Maathius, Eds. IMS Collections, Vol. **9**, 91–104.

[3] Hall, P. and Li, K.-C. (1993). On almost Linearity of Low Dimensional Projections from High Dimensional Data. *The Annals of Statistics* **21**, 2, 867–889.

[4] Leeb, H. (2013). On the Conditional Distributions of Low-Dimensional Projections from High-Dimensional Data. *The Annals of Statistics,* forthcoming.

Score test statistic for change-point detection in AR time series with dependent errors

Katarína Starinská[1]

[1] *Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague*

Abstract

Detecting changes in the parameter values of any model is of great importance for many sectors. With a model up to date we are able to give better predictions. Finding a change-point can help us understand the influence of some events on observed data. The efficient score test statistic was introduced in [5] for detecting changes in the parameters of autoregressive(AR) time series with independent identically distributed(i.i.d) errors. This test allows us to detect a change in all the parameters at once or in every parameter separately. We study the behavior of this statistic when the assumption of i.i.d. white noise is violated and replaced with the assumption of having martingale difference sequence. We present the simulation study which shows us the asymptotic behavior and the strength of this test statistic.

Keywords: Change-point detection, Invariance principle, Autoregressive time series.
AMS subject classifications: primary 62F05, secondary 60F17, 62M10

Bibliography

[1] Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*, Chichester: Wiley.

[2] Davidson, J. (1994). *Stochastic Limit Theory: Advanced Texts in Econometrics*, Oxford University Press, USA.

[3] Davis, R.A., Huang, D. and Yao, Y.-C. (1995). Testing for a Change in the Parameter Value and Order of an Autoregressive Model. *Ann. Statist.*, 23, 282-304.,

[4] Eberlein, E. (1986). On Strong Invariance Principles Under Dependence Assumptions. *Ann. Probab.*, 14, 260-270.

[5] Gombay, E. (2008). Change Detection in Autoregressive Time Series. *J. Multivar. Anal.*, 99, 451-464.

[6] Hušková, M., Prášková, Z. and Steinebach, J. (2007). On the Detection of Changes in Autoregressive Time Series I & II. *J. Stat. Plann. Inference*, 137.

Stability testing for a special Pickard model

Sándor Baran[1], **Gyula Pap**[2] **and Kinga Sikolya**[1]

[1]*Faculty of Informatics, University of Debrecen*
Kassai út 26, H–4028 Debrecen, Hungary
[2]*Bolyai Institute, University of Szeged*
Aradi vértanúk tere 1, H-6720 Szeged, Hungary

Abstract

We investigate the least squares estimator of the stability parameter $\varrho := |\alpha| + |\beta|$ for a spatial unilateral autoregressive process

$$X_{k,\ell} = \alpha X_{k-1,\ell} + \beta X_{k,\ell-1} + \varepsilon_{k,\ell},$$

where the independent innovations $\varepsilon_{k,\ell}$ have zero mean and unit variance. In the unstable case $\varrho = 1$ [see, [1]] we show the asymptotic normality of the estimator with a scaling factor $n^{5/4}$, in contrast to the classical AR(p) model, where the least squares estimator of the appropriate stability parameter is not asymptotically normal. The limiting distribution of the stability parameter can be applied for building unit root tests for the above spatial process. In the unstable case we also obtain the limiting distribution of the least squares estimator of the parameters (α, β) and give an essentially simpler proof of the corresponding results of [2].

Keywords: unstable spatial unilateral autoregressive process, unit root tests.
AMS subject classifications: primary 62M10; secondary 62F12
Acknowledgements: This research has been supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

Bibliography

[1] Basu, S. and Reinsel, G. C. (1993) Properties of the spatial unilateral first-order ARMA model, *Adv. Appl. Prob.* **25**, 631–648.

[2] Baran, S., Pap, G., Zuijlen, M. v. (2007) Asymptotic inference for unit roots in spatial triangular autoregression. *Acta Appl. Math.* **96**, 17–42.

A Bayesian approach for the joint analysis of fMRI and SNP data

Francesco C. Stingo[1], Michele Guindani[1], Marina Vannucci[2] and Vince D. Calhoun[3]

[1] *Department of Biostatistics, University of Texas MD Anderson Cancer Center,*
[2] *Department of Statistics, Rice University,*
[3] *Depts. of Electrical and Computer Engineering, University of New Mexico*

Abstract

In this work we develop a Bayesian hierarchical modeling approach for imaging genetics. We have available data from a study on schizophrenia. Our interest lies in identifying brain regions of interest (ROIs) with discriminating activation patterns between schizophrenic and control subjects, and in relating the ROIs activations with available genetic information from single nucleotide polymorphisms (SNPs) on the subjects. For this task we develop a hierarchical mixture model that includes several innovative characteristics: first, it incorporates the selection of features (ROIs) that discriminate the subjects into separate components, allowing for a direct assessment of the uncertainties in the estimates of the model selection parameters. Second, it allows the mixture components to depend on selected covariates, i.e., the SNPs data. In this sense, our proposed model is integrative, in that it combines the activation patterns with the subjects specific genetic information. Third, it incorporates prior knowledge via network models that capture structural dependencies among the variables. More specifically, it employs spatially based selection process priors that capture available knowledge on connectivity among regions of the brain, so that regions belonging to same activation patterns are more likely to be selected together. Furthermore, our hierarchical formulation accounts for additional spatial correlation among selected features. Applied to the schizophrenia data, the model leads to the simultaneous selection of a set of discriminatory ROIs and the relevant SNPs, together with the reconstruction of the correlation structure of the selected regions. To the best of our knowledge, our work represents the first attempt at a rigorous modeling strategy for imaging genetics data that incorporates all such features.

Keywords: Bayesian Hierarchical Model, fMRI, Imaging Genetic, Markov Random Field, Variable Selection
AMS subject classifications: 62H30, 62H35.

Model selection approach for genome wide association studies in admixed populations

Piotr Szulc[1]

[1]*Department of Mathematics and Computer Science*
Wroclaw University of Technology, Poland

Abstract

The main purpose of Genome Wide Association Studies (GWAS) is the identification of genes responsible for quantitative traits (Quantitative Trait Loci, QTL) or disease causing genes in human populations. Localization of genes in such outbred populations is relatively difficult. The problem comes from the fact that due to cross-overs (exchange of genetic material among chromatids), which occur during production of reproductive cells, the statistical relation between a QTL genotype and a genotype of the neighboring marker might be very small. Therefore the scientists need to use a huge number of densely spaced markers, which necessitates the application of stringent multiple testing corrections and results in a relatively low power of gene detection.
Modifications of Bayesian Information Criterion, mBIC and mBIC2, were successfully used in GWAS and results can be found *e.g.* in [1]. However, it turns out that we can find much more influential genes if we perform GWAS in *admixed population*, obtained as a result of interbreeding between previously separated ancestral populations. In that case, apart from genotypes, we have information about the origin of genome's fragments.
I will present the problem of localizing genes and show how to use modifications of model selection criteria in admixed population. Finally, I will present results of simulations.

Keywords: linear regression, model selection criteria, GWAS, admixed population
AMS subject classifications: 62J05, 92D20

Bibliography

[1] F. Frommlet, F. Ruhaltinger, P. Twarog and M. Bogdan (2011). A model selection approach to genome wide association studies. *Computational Statistics and Data Analysis* 56, 1038–1051.

Modeling the link between two genes expressions and the toxicity of some chemotherapy drugs for chronotherapy of cancer

Mircea Dumitru[1], **Ali Mohammad-Djafari**[1] **and Francis Lévi**[2], **XiaoMei Li**[2]

[1]*Laboratoire des signaux et systems (L2S), SUPELEC-Université Paris Sud*
[2]*Rythmes Biologiques et Cancers (RBC), INSERM - Université Paris Sud*

Abstract

In this article, we present the mathematical model for predicting the optimal circadian timing of an anticancer drug, irinotecan using the circadian expression of two clock genes as input data. Two main classes are represented by those gene expressions signals: mutant and non-mutant mice. The data representing the input are sampled at every three hours along the twenty four hour scale , in the non-mutant case, and at every four hours in the mutant case. The data representing the output (the Body Weight Loss signal) is sampled at every four hours. Both input and output data represent the mean values of the measured data at every point. The proposed model is a linear model using a Maximum A Posteriori Bayesian inference method. The model is first implemented on mean signals, the prediction matrix being built on a certain number of available signals, the rest of signals being used for checking the accuracy of prediction. Because the number of cases in some classes is very law, another implementation (model) using the individual signals is used. When using the individual data, we study the number of training data on the accuracy of the predictions. Two measures, L_1 and L_2 relative distances are used for measuring quality of the global shape of prediction and number of times the position minimum values correspond (since the minimum value corresponds to the optimal drug administration).

Keywords: Circadian Time, Bayesian Maximum A Posteriori (BMAP)
AMS subject classifications: 62F15
Acknowledgements: This work has been partially founded by the C5Sys project (Circadian and Cell cycle Clock systems in Cancer) of EraSysBIO+

Bibliography

[1] Mohammad-Djafari, A. (2012). Bayesian approach with prior models which enforce sparsity in signal and image processing. *EURASIP Journal on Advances in Signal Processing, Special issue on Sparse Signal Processing* 2012-52

Statistical methods for RNAseq coverage profiles and their applications in biological phenomena.

Alicja Szabelska[1,2], Idzi Siatkowski[1], Michal Okoniewski[2]

[1] *Poznan University of Life Sciences, Wojska Polskiego 28, 60-637 Poznan*
[2] *Functional Genomics Center Zurich, Winterthurerstrasse 190, 8104 Zurich*

Abstract

Nowadays next generation sequencing, mostly RNA-Seq is one of the most powerful tools that can be applied to gain more insight in genomic issues. RNA-Seq produces the enormous amounts of data. To be able to find interesting features in such data the most popular parametric approaches (see [1], [5]) summarize the data to "count data" on the gene level and apply negative binomial distribution to model it with usage of generalized linear model. Other popular non-parametric [3] or based on empirical Bayes [2] methods were also introduced to analyze RNA-Seq data. However, they are also based on count data, which means the initial information produced by sequencer is reduced and we do not take advantage of the power of next-generation sequencing tools. That is why we present the methods for seeking and statistical verification of the genomic features based on coverage profiles of interesting genome. According to [4] the first step of the analysis is to introduce a measures of dissimilarity of the profiles. The second step is the statistical validation of the outcomes with usage of permutation test.
During the talk the overall work for coverage based approach will be presented. In addition, the examples of usage of the method in biological experiment will be included.

Keywords: Differential expression, alternative splicing, RNA-Seq, coverage profiles, permutation tests
AMS subject classifications: 62P10, 92B15
Acknowledgements: Research supported by Poznan University of Life Sciences and Functional Genomics Center Zurich (Scientific Exchange Programme NMS^{ch} - no. 11.182)

Bibliography

[1] Anders, S., & Huber, W. (2010). Diferential expression analysis for sequence count data. *Genome Biology*, 11: R106.
[2] Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: empirical Bayesian methods for identifying diferential expression in sequence count data. *BMC bioinformatics*, 11(1), 422.
[3] Li, J., & Tibshirani, R. (2011). Finding consistent patterns: A nonparametric approach for identifying diferential expression in RNA-Seq data. Statistical Methods in Medical Research.

[4] Okoniewski, M. J., Leniewska, A., Szabelska, A., Zyprych-Walczak, J., Ryan, M., Wachtel, M., Morzy, T., Schaer, B., & Schlapbach, R. (2012). Preferred analysis methods for single genomic regions in RNA sequencing revealed by processing the shape of coverage. Nucleic Acids Research, 40(9), e63-e63.
[5] Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a bioconductor package for diferential expression analysis of digital gene expression data. Bioinformatics, 26: 139140.

Statistical methodology in the scope of performance budgeting

Žiga Kotnik[1] **and Maja Klun**[1]

[1]*Faculty of Administration, University of Ljubljana*

Abstract

The environmental protection has become one of the main political priorities of the United Nations and the European Union. Environment is one of the areas where measurement of performance and efficiency is particularly difficult specially owing to lack of information and absence of traceability of actual effects on the environment. For this reason, environment requires its own approach that will properly evaluate environmental data and use them when planning the budget. Performance budgeting promises such solution as this approach investigates the linkage between spent public resources and planned public policy objectives. Realization of these objectives is measured through a set of indicators, attributed to each objective. The purpose of this paper is to present a broader theoretical and methodological framework of performance budgeting in the field of environment and set a proper environmental model for studying the linkage between environmental taxes, environmental expenditures and environmental impacts that are all interrelated. These will be estimated by a specifically tailored statistical model and tested in the case of the EU Member States.

Keywords: Performance budgeting, statistical methodology, the European Union, environment.
AMS subject classifications: 62M10.

Stochastic interest rates in life insurance mathematics

Gábor Szűcs[1]

[1] *Comenius University; Faculty of Mathematics, Physics and Informatics; Department of Applied Mathematics and Statistics*

Abstract

Basic life insurance mathematics applies some simplifications, e.g., the assumption of constant interest rates during the period of insurance (see [1]). Insurance companies in Slovakia usually follow this assumptions and they calculate the premium using the technical interest rate. According to Decree of the National Bank of Slovakia the maximum technical rate of interest shall be 2.5% p. a. From a practical point of view, insurance corporations invest collected premiums on behalf of policyholders in different types of assets (e.g., bonds, shares, deposits). However, their yields have stochastic character, because the situation on the financial and capital markets is continually changing. For insurance companies it is important to know what kind of risks and losses will they face, if premium is computed using technical interest rate, while return on investments isn't guaranteed. The aim of this paper is to present several methods for pricing the present value of potential future insurance losses. We assume that the potential losses are derived from the stochastic behavior of interest rates and market yields.

Keywords: technical rate of interest, actuarial present value, Vasicek model, ARIMA-process
AMS subject classifications: 91B30, 91G30
Acknowledgements: This research was supported by the Slovak grant VEGA No. 2/0038/12.

Bibliography

[1] Gerber, H. U. (1997). *Life Insurance Mathematics*, Third Edition, Springer-Verlag Berlin Heidelberg. ISBN 3-540-58858-3.

Analysis of the approximate maximum likelihood estimators of diffusion parameters by simulations

Snježana Lubura[1] **and Miljenko Huzak**[1]

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

Abstract

Observations of diffusion paths are usually discrete. Except in few cases, the exact maximum likelihood estimation (MLE) of diffusion parameters is not possible. Hence, some other methods of estimation have to be applied. We analyze some of the approximate maximum likelihood estimators by using computer simulations. Diffusion models we used are growth models such as Gompertz, logistic and von Bertalanffy diffusion models.

Keywords: approximate MLE, simulation, diffusion processes.
AMS subject classifications: 60J60, 60J70, 62F10, 62F12, 62M05, 65C30.

Bibliography

[1] Huzak, M. (2001). A general theorem on approximate maximum likelihood estimation. *Glasnik Matematički* Vol. 36 (56), 139–153.

[2] Huzak, M. (1998). Parameter estimation of diffusion growth models. *Mathematical Communications* 3, 129–134 221–225.

[3] Kloeden, P.E., Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag.

[4] Prakasa Rao, B.L.S. (1999). *Statistical inference for Diffusion Type Processes*, Arnold, London.

Adaptive estimation in mixture models with varying mixing probabilities

A.V. Doronin[1]

[1] *Kyiv National Taras Shevchenko University, Kyiv, Ukraine*

Abstract

Semiparametric estimation problems are considered for a model of finite mixture with mixing probabilities varying from observation to observation. We present estimators based on adaptive estimating equations, and compare them with estimators of two another types, namely the moment and quantile ones. Performance of these estimators is compared both analytically and by simulations.

Keywords: Finite mixture model, adaptive estimation, simulation, generalized estimating equation.
AMS subject classifications: 62F12, 62F35, 62G05, 62G35, 62G20.

Bibliography

[1] Maiboroda, R.E., Sugakova, O.V. and Doronin, A.V. Generalized estimating equations for mixtures with varying concentrations. *The Canadian Journal of Statistics* to appear. Published on-line http://onlinelibrary.wiley.com/doi/10.1002/cjs.11170/abstract.
[2] Doronin, A.V. (2012). Robust Estimates for Mixtures with Gaussian Component. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics & Mathematics* (in Ukrainian). 1, 18–23.
[3] Maiboroda, R.E. and Kubaichuk, O.O. (2005). Improved estimators for moments constructed from observations of a mixture. *Theory of Probability and Mathematical Statistics*. 70, 83–92.
[4] Maiboroda, R.E. and Sugakova, O.V. (2008). *Estimation and classification by observations from mixtures*. Kyiv University Publishers, Kyiv (in Ukrainian).

Models for dependent paired comparison data

Manuela Cattelan[1] **and Cristiano Varin**[2]

[1] *Department of Statistical Sciences, University of Padova*
[2] *D.A.I.S., Ca' Foscari University - Venice*

Abstract

Paired comparison data are binary data that originate from the comparison of objects in couples. This type of data arises in many applications including animal behaviour experiments, sport tournaments and psychometric experiments. The aim of the analysis is to determine whether some covariates can predict the results of comparisons or to estimate a "worth" parameter for each of the elements compared which can be used for ranking them.
The Bradley-Terry and the Thurstone models are usually employed to analyse paired comparison data. The Bradley-Terry model is a logistic model in which the linear predictor is described as the difference of the worth parameters of the two elements compared. The Thurstone model is similarly specified, but it employs a probit link function. However, usually both models are estimated assuming that all comparisons are independent. This assumption is often unrealistic, for example in animal behaviour experiments it seems inappropriate to assume that the results of two different fights involving a common animal are independent. Extensions of traditional models that overcome this limit are proposed. Unfortunately, the inclusion of dependence generates some difficulties in the estimation of the model since the dependence among comparisons involving the same element produces an intricate scheme of cross-correlations among observations and sometimes, as in animal behaviour experiments, it is not possible to distinguish groups of independent data. The problem of estimating the extension of the Bradley-Terry model for dependent data is overcome by resorting to optimal estimating equations. The dependence structure is taken into account in the covariance matrix of the observations which includes parameters that cannot be estimated independently of the regression parameters. This problem is solved by means of a hybrid pairwise likelihood estimation which cycles between optimal estimating equations for estimation of the regression parameters given the covariance parameters, and maximum pairwise likelihood for estimation of the covariance parameters given the regression parameters. On the contrary, the extension of the Thurstone model can be seen as a multivariate probit model, so many different estimation techniques can be employed. The properties of different estimation methods in models for dependent paired comparison data will be discussed and illustrations to real data applications will be presented.

Keywords: Optimal estimating equations, Paired comparisons, Pairwise likelihood.
AMS subject classifications: 62F10, 62J12.

Adaptive Bayesian estimation in Gaussian sequence space models

Jan Johannes[1], **Rudolf Schenk**[1] **and Anna Simoni**[2]

[1] *Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve*
[2] *Université de Cergy-Pontoise, 33 boulevard du Port, F - 95011 Cergy-Pontoise*

Abstract

We consider the inverse regression problem $Y = Af + \sqrt{\varepsilon}\xi$, where A is a known linear operator between two Hilbert spaces, ξ a Gaussian white noise, and ε a noise level. The objective of this paper is to establish adaptive nonparametric posterior concentration rates of convergence for the regression function f. In a first step, we derive lower and upper bounds for the posterior concentration rates over a family of Gaussian prior distributions indexed by a tuning parameter. These rates are based on tail bounds for noncentral χ^2 distributions established in [1]. By selecting the optimal tuning parameter over the class we derive the fastest bounds within the family. Of course, this optimization procedure depends on the regression function f and leads to an oracle in the frequentist framework. In a second step, we put a prior on the tuning parameter and derive the posterior concentration rate of the constructed hierarchical Gaussian prior distribution. The results of the Bayesian inference are furthermore put into relation with the frequentist problem of estimating the regression function f. Considering typical smoothness classes, we show that a full data-driven Bayes estimate derived from this hierarchical prior distribution can attain minimax optimal rates of convergence and is hence adaptive.

Keywords: adaptive estimation, Bayesian inference, Gaussian sequence, minimax theory, nonparametric estimation.
AMS subject classifications: 62F15.
Acknowledgements: Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged.

Bibliography

[1] L. Birgé. (2001). An alternative point of view on Lepskis method. *IMS Lecture Notes, State of the art in probability and statistics, (Leiden 1999).* 36, 113–133.

Stochastic versions of logarithmic Sobolev inequalities

Aygul Abakirova[1]

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

Abstract

Our main results refer to the stochastic versions of the well-known Poincare and logarithmic Sobolev inequalities. The Poincare inequality for Gaussian variables was formulated in connection with classical isoperimetric problem. The log-Sobolev inequality was established by Gross in 1975 as an equivalent condition for hypercontractivity of associated Markov semigroup. This work gave start for the further research. Log-Sobolev inequalities are closely linked with transport and entropy-information inequalities etc, they were intensively studied in probability theory, geometry, statistical mechanics [2], [3].
We extend the results from Gaussian case to infinitely divisible variables with indication of optimal constants. Method applied is based on the idea of "embedding" such variables in infinitely divisible process, its dynamics allows us to use Ito's formula and Kolmogorov equations. Estimates are obtained in terms of triplet of local characteristics. The principal advantage of log-Sobolev inequalities is dimension independence (for example, standard Sobolev inequalities are not dimension independent). The method of proof based on stochastic calculus and integral representations works naturally on path spaces [1].
We obtain the Poincare and the log-Sobolev inequality for the skew Brownian motion. The skew Brownian motion is a unique strong solution of a stochastic equation $X_t = B_t + (2\alpha - 1)L_t^0(X), t \geq 0, (\alpha \in [0, 1])$, where $L^0(X)$ is the local time of the unknown process X at zero. Such generalized diffusion can be used to model permeable barrier. Drift coefficient is not defined at zero and this effect is translated into the backward equation as a transmission condition, we have to apply generalized change of variable (Ito-Tanaka) formula etc. The estimates depend on the local time of the process. Different points of view on skew Brownian motion, its applications and generalizations can be found in Lejay[4].
Effectiveness of log-Sobolev inequalities in infinite dimensional analysis is illustrated in statistical mechanics and analysis on path spaces. Different versions of log-Sobolev inequalities are appeared to be useful in probability, differential equations, combinatorics. Being a standard application of log-Sobolev type inequalities concentration of measure phenomenon is useful in statistics as well as in geometry.

Keywords: Poincaré inequality, logarithmic Sobolev inequality, skew Brownian motion, Malliavin calculus, Concentration of measure.
AMS subject classifications: 60E15, 60J60, 60H07.

Bibliography

[1] Abakirova, A. (2012). Functional inequalities on pass spaces of processes with independent increments. *Russian Math. Surveys* 67:2, 381–383.

[2] Ane, C., Blachere, S., Chafaï, D., Fougeres, P., Gentil, I., Malrieu, F., Roberto, C. and Scheffer, G. (2000). *Sur les inegalites de Sobolev logarithmiques*. Panoramas et Syntheses 10, Soc. Math. de France.
[3] Guionnet, A. and Zegarlinski, B. (2003). *Lectures on logarithmic Sobolev inequalities*, Seminaire de Probabilites XXXVI, Lecture Notes in Math., vol. 1801, Springer, Berlin, 1–134.
[4] Lejay, A. (2006). On the constructions of the skew Brownian motion. *Probab. Surveys* 3, 413–466.

On oracle inequality for exponential weighting of ordered smoothers

Chernousova, E.[1], Golubev, Yu.[2,3] and Krymova, E.[1,2,4]

[1] *Moscow Institute of Physics and Technology (State University),*
[2] *Institute for Information Transmission Problems,*
[3] *CNRS, Université de Provence,*
[4] *DATADVANCE*

Abstract

This paper deals with recovering an unknown vector from noisy data with the help of special family of linear estimates, namely, a family of ordered smoothers. The estimators within this family are aggregated using the exponential weighting method. Our goal is to derive oracle inequalities controlling the risk of the aggregated estimate. Based on probabilistic properties of the unbiased risk estimate, we propose a new method for obtaining oracle inequalities and show that for the exponential weighting we can get better remainder terms than the one in Kneip's oracle inequality [1].

Keywords: ordered smoother, exponential weighting, unbiased risk estimation, oracle inequality
AMS subject classifications: 62G05
Acknowledgements: This research was supported by Laboratory of Structural Methods of Predictive Modeling and Optimization, Moscow Institute Physics and Technology, Russian Federation government grant, ag. 11.G34.31.0073.

Bibliography

[1] Kneip, A. (1994). Ordered linear smoothers. *Annals of Stat.* 22, 835–866.
[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle *Proc. 2nd Intern. Symp. Inf. Theory.* 267–281.
[3] Golubev, Yu. (2012). Exponential weighting and oracle inequalities for projection methods. *Problems of Information Transmission* no. 3. *arXiv:1206.4285*

Upper and lower bounds for ordered random variables

Nuria Torrado[1]

[1]*Department of Statistical Methods, Universidad de Zaragoza, Ciudad Escolar s/n, 44003 Teruel (Spain)*

Abstract

Our aim was to examine upper and lower bounds for some reliability functions for independent but not identically distributed random variables. This problem was studied by different authors when the random variables are independent and identically distributed (see [2, 3, 4], among others).
In the article and in the presentation a short overview on the wide field of stochastic orderings is given, showing some results given by Torrado and Lillo [5] and also some of the current research the author is doing at the moment. Some applications to multiple-outlier models will be briefly discussed. Multiple-outlier models are interesting due to applications in the study of the robustness of different estimators of parameters of a wide range of distributions, see e.g. Balakrishnan [1].

Keywords: reliability theory, multiple-outlier models, ordered random variables, stochastic orderings.
AMS subject classifications: 60E15, 60K10, 62G30.
Acknowledgements: The financial support of the Spanish Ministry of Education and Science under grant SEJ2007-64500 is acknowledged.

Bibliography

[1] Balakrishnan, N. (2007). Permanents, order statistics, outliers, and robustness. *Revista Matemática Complutense* 20, 7–107.
[2] Kochar, S.C. and Korwar, R. (1996). Stochastic orders for spacings of heterogeneous exponential random variables. *Journal of Multivariate Analysis* 57, 69–83.
[3] Kochar, S.C. and Xu, M. (2011). Stochastic comparisons of spacings from heterogeneous samples in Advances. In M. Wells and A. Sengupta, editors, Advances in Directional and Linear Statistics, pp. 113–129. Festschrift Volume for J.S. Rao, Springer.
[4] Pledger, G. and Proschan, F. (1971). Comparisons of order statistics from heterogeneous populations, with applications in reliability. In: J.S. Rustagi Ed., Optimizing Methods in Statistics, Academic Press, New York, p.p. 89–113.
[5] Torrado, N. and Lillo, R.E. (2013). Likelihood ratio order of spacings from two heterogeneous samples. *Journal of Multivariate Analysis* 114, 338–348.

Mixture Lorenz curves. Three new models

Drăgulin Mircea[1] **and Gheorghe Carmen Adriana**[2]

[1]*Faculty of Mathematics and Computer Science, University of Bucharest*
[2]*National Institute of Economic Research, Romanian Science Academy*

Abstract

The Lorenz curve is one of the most investigated and also significant tool in the study of distribution and inequality of income. The main difficulty in finding a good analytical form is the lack of fitting on some intervals, especially in the tail of the function. Mixture parametric approach may overdue these problematical issues by introducing better constraints.
In this paper, three new mixture Lorenz Curves are generated from initial Lorenz Curve families. In order to analyze the inequality in the income distribution, for the third proposed curve the Gini indexes are obtained.

Keywords: parametric Lorenz curve, Gini index.
AMS subject classifications: 60E15, 91B82
Acknowledgements: We would like to thank to our supervisor of this project, professor Vasile Preda for the valuable guidance and advice.

Bibliography

[1] Sarabia, J.M.; Castillo, E.; Pascual, M; Sarabia, M. (2005). Mixture Lorenz curves. *Economics Letters,* Elsevier, vol. 89(1).
[2] Wang, Z. X.; Ng, Y-K.; Smyth, R. (2007). Revisiting the ordered family of Lorenz curves. *Discussion paper 19/07* Department of Economics Monash University.

Bayesian multiscale analysis of images

Leena Pasanen[1] and Lasse Holmström[1]

[1]*Department of Mathematical Sciences, University of Oulu, Finland*

Abstract

Two novel multiscale methods for digital images are proposed. The first method detects differences between two images obtained from the same object at two different instants of time. It detects both small scale, sharp changes and large scale, average changes. The second method extracts features that differ in intensity from their surroundings and produces a multiresolution analysis of an image as a sum of scale-dependent components.

As images are usually noisy, Bayesian inference is used to separate real differences and features from artefacts caused by random noise. The use of the Bayesian paradigm allows the use of various noise types, incorporation of expert knowledge about the images at hand and facilitates analysis of non-linear transformation of images.

The methods are instants of SiZer (Significant zero crossings of derivatives) methodology that was originally considered for one-dimensional nonparametric probability density estimation and curve fitting [1, 2]. The new methods, iBSiZer (Bayesian SiZer for images) and MRBSiZer (Multiresolution Bayesian SiZer), were originally proposed in [3] and [4], respectively.

Keywords: Bayesian methods, Scale space, Image analysis, SiZer
AMS subject classifications: 62M40

Bibliography

[1] Chaudhuri, P. and Marron, J. S. (1999). SiZer for Exploration of Structures in Curves. *Journal of the American Statistical Association* 94, 807-823.

[2] Chaudhuri, P. and Marron, J. S. (2000). Scale Space View of Curve Estimation. *The Annals of Statistics* 28, 408428.

[3] Holmström, L. and Pasanen, L. (2012). Bayesian Scale Space Analysis of Differences in Images. *Technometrics* 54, 16-29.

[4] Holmström, L., Pasanen, L., Furrer, R. and Sain, S. R. (2011). Scale Space Multiresolution Analysis of Random Signals. *Computational Statistics & Data Analysis* 55, 2840-2855.

The comparison of Harris corner detection method and statistical methods for detection of dominant points of two-dimensional objects

Çiğdem Güngör[1] **and Orhan Kesemen**[2]

[1] *Department of Statistics, Middle East Technical University, Ankara, Turkey*
[2] *Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, Turkey*

Abstract

Dominant point detection problem is one of the most important issues of image analysis researches since dominant points of an object carry the most important information within the boundary points of that object. The main aim of researching the dominant points of an object is to be able to represent an object on the digital image with a limited number of points. We need such a simplification to reduce the unnecessary large amount of raw data. The object detection algorithms work efficiently. The Harris Corner Detection, which is fundamentally based on thresholding technique, is a well-known method which is used for detection the dominant points of 2-dimensional objects. In this paper we compare the performances of the Harris Corner Detection algorithm and statistical methods that are local and global which I proposed as a master thesis in 2012. This study aims to determine the strengths of both approaches relative to each other and open a way for further studies such as optimal number of dominant points of an object in a digital image.

Keywords: Dominant point detection, pattern recognition, corner detection
AMS subject classifications: 68T10

Bibliography

[1] Güngör, Ç. (2012). *Optimization Algorithms For Detection Of Dominant Points Of Objects In Binary Images*, Master Thesis, Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, Turkey.
[2] Harris, C. and Stephens, M. (1988). *A Combined Corner and Edge Detector*, Proceedings of The Fourth Alvey Vision Conference, Manchester, UK, 147–151.
[3] Orguner, U. and Gustafsson, F. (2007). *Statistical Characteristics Of Harris Corner Detector*, SSP '07 Proceedings of the 2007 IEEE/SP 14th Workshop on Statistical Signal Processing, IEEE Computer Society Washington DC, USA, 571–575.

Multiresolution methods for ranking

Eric Sibony[1]

[1]*Institut Mines-Telecom, LTCI, Telecom Paristech / CNRS, France*

Abstract

We use a recently introduced framework for multiresolution analysis on the symmetric group to predict rankings. Viewing preferences as sets of permutations, ranking prediction implies to handle probability distributions on the symmetric group, which usually leads to intractable storage or computations. We define a new smoothing technique based on wavelet decomposition that allows to obtain sparse representations for a large class of probability distributions. We show that in many practical cases, our method performs efficiently, in terms of storage and from a computational cost perspective as well.

Keywords: Ranking, Multiresolution analysis, Wavelets, Statistical estimation, Pairwise preferences.
AMS subject classifications: 62G05, 43A65.
Acknowledgements: This work is supported by the Labex LMH grant n°ANR-11-IDEX-0003-02.

Bibliography

[1] Diaconis, P. Group representations in probability and statistics. (1988). *Institute of Mathematical Statistics Lecture Notes - Monograph Series.*
[2] Jagabathula, S and Shah, D. (2011). Inferring Rankings Using Constrained Sensing. *IEEE Transactions on Information Theory* 57(11):7288–7306.
[3] Kondor, R and Dempsey, W. (2012). Multiresolution analysis on the symmetric group. *Neural Information Processing Systems* 25.
[4] Stéphane Mallat. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE* II(7).
[5] Mingxuan Sun, M., Lebanon, G. and Kidwell, P. (2012). Estimating probabilities in recommendation systems. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61(3):471–492.

Uniform-in-bandwidth kernel estimation for censored data

Sarah Ouadah[1]

[1]*Laboratoire de Modélisation Aléatoire, Université Paris Ouest Nanterre la Défense–Paris 10, France*
Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie– Paris 6, France

Abstract

We present limit laws for the nonparametric kernel lifetime density and hazard rate estimators in a right random censorship model. These limit laws are uniform with respect to the choices of bandwidth and kernel and are established in the framework of convergence in probability. *Uniform-in-bandwidth* results are useful to describe the limiting behavior of kernel estimators with random or data-dependent bandwidths. Furthermore, we allow the bandwidth to vary within the complete range for which the estimators are consistent and we provide explicit values for the asymptotic limiting constant for the sup-norm of the estimation random error.

Keywords: Functional limit laws, Nonparametric kernel density estimation, Right random censorship model, Hazard rate function, Kaplan-Meier empirical process.
AMS subject classifications: 62G07, 62G20, 62N01, 62N02, 60F17.

Bi-log-concave distribution functions and confidence bands

Petro Kolesnyk[1]

[1]*Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland*

Abstract

In nonparametric statistics one is often interested in estimators or confidence regions for curves such as densities or regression functions. Estimation of such curves is typically an ill-posed problem and requires additional assumptions. An interesting alternative to smoothness assumptions are qualitative constraints, e.g. monotonicity, concavity or log-concavity. Estimation of a distribution function F based on independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ with c.d.f. F is less difficult. But non-trivial confidence regions for certain functionals of F such as the mean do not exist without substantial additional constraints (Bahadur and Savage, 1956).
In density estimation, a particular constraint which attracted considerable attention recently is log-concavity. That means, we estimate a probability density f on $\mathbb{R}^d$ under the constraint that $\log f : \mathbb{R}^d \to [-\infty, \infty)$ is a concave function. While many papers are focussing on point estimation, Schuhmacher et al. (2011) show that combining the log-concavity constraint and a standard Kolmogorov-Smirnov confidence region yields an interesting nonparametric confidence region, although its explicit computation is far from obvious. In the present work we introduce a new and weaker constraint on distribution functions:
A distribution function F on the real line is called *bi-log-concave* if both $\log F$ and $\log(1 - F)$ are concave functions (with values in $[-\infty, 0]$).
This new shape constraint is rather natural in many situations. For instance, any c.d.f. F with log-concave density $f = F'$ is bi-log-concave, according to Bagnoli and Bergstrom (2005). But bi-log-concavity of F alone is a much weaker constraint: F may have a density with an arbitrarily large number of modes. Various characterizations of bi-log-concavity are provided. It is shown that combining any nonparametric confidence band for F with the new shape-constraint leads to substantial improvements and implies non-trivial confidence bounds for arbitrary moments and the moment generating function of F.

Keywords: Shape constraints, log-concavity, confidence set, Empirical distribution, Kolmogorov-Smirnov.
AMS subject classifications: 62G05, 62G15, 62G20, 62G30
Acknowledgements: This is joint work with Lutz Dümbgen (Bern) and Ralf Wilke (Nottingham).

Bibliography

[1] Owen, A. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* ***90(430)***, 516-521.
[2] Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Econ. Theory* ***26***, 445-469.
[3] Bahadur, R. and Savage, L. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* ***27***, 1115-1122.
[4] Schuhmacher, D., Hüsler, A. and Dümbgen, L. (2011). Log-concave distributions as a nearly parametric model. *Statist. Risk Model.* ***28(3)***, 277-295.

Adaptive confidence sets from Bayes procedures

Botond Szabo[1], Aad van der Vaart[2] and Harry van Zanten[3]

[1]*Eindhoven University of Technology*
[2]*Leiden University*
[3]*University of Amsterdam*

Abstract

Adaptive techniques for nonparametric estimation have been widely studied in the literature and many rate-adaptive results have been provided for a variety of statistical problems. However an adaptive estimator without any knowledge of its uncertainty is rather uninformative, since one knows that the estimator is optimally close to the true function, but has no information about the actual distance.
In the Bayesian framework credible sets can be constructed to quantify the uncertainty in the posterior distribution. Due to the recent developments of Bayesian computational methods constructing credible sets can be easier than constructing confidence sets from frequentist estimators. The frequentist coverage of credible sets describes to what extent credible sets can be viewed as frequentist confidence sets.
We consider the problem of constructing Bayesian based confidence sets that are adaptive in L^2-loss over a continuous scale of Sobolev classes in the Gaussian White noise model. We show that both the hierarchical Bayes and marginal likelihood empirical Bayes approaches lead to credible sets with asymptotic coverage zero for certain oddly behaving functions. Then we give a new empirical Bayes method based on the results of [1], which solves this problem and provides uniform and adaptive confidence sets over a whole collection of Sobolev classes.

Keywords: Nonparametric Bayesian procedure, adaptation, credible sets, coverage, Gaussian processes.
AMS subject classifications: Primary 62G15; secondary 62G20.

Bibliography

[1] Robins, James and van der Vaart, Aad W. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* 34, 229–253.

Approximating the posterior distribution of mixture weights with application to transcript expression estimation

Panagiotis Papastamoulis[1] **and Magnus Rattray**[1]

[1]*Faculty of Life Sciences, University of Manchester*

Abstract

This study focuses on approximating the posterior distribution of mixture weights ($\boldsymbol{\theta}$) given some data ($\boldsymbol{x}$) using Variational Bayes (VB) methods [1]. Standard VB implementation [4] for this problem approximates the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{x})$ of parameters and latent variables ($\boldsymbol{z}$). It is demonstrated via simulation that this approach leads to variance underestimation. For this reason a new variational scheme is developed by integrating out the latent variables and targeting the marginal posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$. The new approximation belongs to the richer family of Generalized Dirichlet distributions [8], while a stochastic approximation algorithm [6] performs the optimization in the corresponding spaces arising from two different parameterizations. Moreover, it is proven that the new solution leads to a better marginal log-likelihood bound compared to the former.
The method is applied to transcript expression estimation using high throughput sequencing of RNA (RNA-seq) technology. Mixture models are a natural way to deal with such problems, and Gibbs sampling has already been applied [3]. The application of Variational methods to these datasets is novel and leads to encouraging results. Finally, the variational solution is exploited in order to improve Markov Chain Monte Carlo (MCMC) sampling with the Delayed Rejection algorithm [7].

Keywords: Kullback-Leibler divergence, marginal likelihood bound, BitSeq, RNA-seq
AMS subject classifications: 62F15, 81T80, 92B15

Bibliography

[1] Bishop, C. (2006). *Pattern Recognition and Machine Learning*, Springer, New York.

[2] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Springer, New York.

[3] Glaus, P., Honkela, A. and Rattray, M. (2012). Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* 28, 1721–1728.
[4] Hensman, J., Rattray, M. and Lawrence, N.D. (2012). Fast Variational Inference in the Conjugate Exponential Family. *Advances in Neural Information Processing Systems*, arXiv:1206.5162v2.
[5] Pan, Q., Shai, O., Lee, L.J., Frey, B.J. et al. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* 40:1413–1415.
[6] Spall, J.C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control* 37, 332–341.
[7] Tierney, L. and Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* 18, 2507–2515.
[8] Wong, T.T. (1998). Generalized Dirichlet distribution in Bayesian analysis. *Applied Mathematics and Computation* 97, 165–181.

Alternative based thresholding for pre-surgical fMRI

Joke Durnez[1], Beatrijs Moerkerke[1], Andreas Bartsch[2] and Thomas E.Nichols[3]

[1] *Department of Data Analysis, Ghent University, Belgium*
[2] *Department of Neuroradiology, University of Heidelberg, Germany*
[3] *Department of Statistics & Warwick Manufacturing Group, University of Warwick, United Kingdom*

Abstract

fMRI is a non-invasive neuroimaging technique that enables to locate important psychological tasks in the brain. The procedure plays a major role in pre-surgical planning for patients with resectable brain lesions such as tumors. fMRI studies can guide resection, thereby preserving vital brain tissue.
For an fMRI data analysis, the brain is divided in more than 100,000 voxels. For each voxel, a statistical test classifies the voxel either as active or as unrelated to the task. In cognitive neurosciences, focus lies on controlling the false positive rate to account for the huge multiple testing problem that arises. However, stringent control of false positives implies an increase of false negatives which can be detrimental in clinical settings where false negatives may lead to surgical resection of vital brain tissue. Consequently, we argue for a testing procedure with a stronger focus on preventing false negatives.

We present a thresholding procedure that incorporates information on both false positives and false negatives. We combine 2 measures of significance for each voxel: a classical p-value which reflects evidence against the null hypothesis of no activation and an alternative p-value which reflects evidence against activation with a pre-specified size. This results in a layered statistical map for the brain. One layer consists of voxels exhibiting strong evidence against the null of no activation while a second layer is formed by voxels for which activation cannot be confidently excluded. The third level then shows voxels for which the presence of activation can be rejected.

Keywords: fMRI, power, false negative errors, multiple testing, pre-surgical fMRI
AMS subject classifications: 62P07

Regression based predictions for irradiation doses

Ferenc Rárosi[1,2], Krisztina Boda[2], Zoltan Varga[3] and Zsuzsanna Kahan[3]

[1] *University of Szeged, Bolyai Institute, Hungary*
[2] *University of Szeged, Department of Medical Physics and Informatics, Hungary*
[3] *University of Szeged, Department of Oncotherapy, Hungary*

Abstract

Radiotherapy is an effective treatment for breast cancer, but it can bring significant late morbidity, particularly to the heart. The radiation dose to the heart may individually vary in the supine versus the prone position in most cases. For the gold standard decision about the preferable treatment position, series of CT scans and therapy plans are needed in both positions. This method is expensive and means extra radioactive dose to the patients.
Our goal was to set up a parsimonious classifier method that predicts the preferable setup.
The dataset consisted of 138 patients of breast cancer measured in both positions. Indicator of the risk is the irradiation dose to the left anterior descendent coronary artery (LAD). The LAD dose difference (D) is a continuous dependent variable. For some classification methods it was dichotomised according to its sign.
We compared different classification methods (multiple logistic regression, discriminant analysis, decision tree, neural networks), finally the use of a multiple linear regression gave the best results with D as dependent variable and three predictors. We estimated not only the proportion of misclassified patients, but also the distribution of the misclassified dose also with a 1000 times random cross validation method. Three SPSS macros (regression macro, recode macro, cleaning macro)

were used. The optimal cutpoint was determined with ROC (receiver operating characteristics) analysis.
The regression based method resulted with a good classification of the cases (overall accuracy: greater than .83, average misclassified doses: less than 2Gy).

Keywords: Linear regression, cross-validation, ROC analysis.
AMS subject classifications: 62J05.

Bibliography

[1] Nancy A. Obuchowski, Michael L. Lieber, and Frank H. Wians: ROC Curves in Clinical Chemistry: Uses, Misuses, and Possible Solutions *Clinical Chemistry* 50, pp 11181125, 2004

[2] Zsuzsanna Kahan, Zoltn Varga, Adrienn Cserhti, Krisztina Boda, Katalin Hideghty, Lszl Thurz: Individualised positioning for breast radiotherapy: a comprehensive approach for OAR protection *Radiotherapy and Oncology* 99, pp. S556-S557, 2011

Competing risks analysis in nephrology research: An example in peritoneal dialysis

Laetitia Teixeira[1] and Denisa Mendonca[2]

[1] *Doctoral Program in Applied Mathematics - Faculty of Sciences and Institute of Biomedical Sciences Abel Salazar, University of Porto*
[2] *Institute of Biomedical Sciences Abel Salazar and Public Health Institute, University of Porto*

Abstract

In clinical and epidemiological research, increasing importance has been given to the competing risk approach and this methodology has been referred as the rule rather than the exception in follow-up studies [1]. It is an extension of classical survival analysis.
In the competing risks framework, patients may fail to one of the K possible causes. A competing risk is an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event [2].
In the presence of competing risks, two types of analysis can be performed: modelling the cause-specific hazard and modelling the hazard of the subdistribution [3,4]. The context of the research question is the main determinant for the choice of an appropriate statistical model. When the hazard of the subdistribution is analyzed, the goal is to compare the probability of the event of interest and therefore

can be translated into actual numbers of patients with this event. Comparing the cause-specific hazards gives an insight into the biological process [3,4,5].
In peritoneal dialysis programs, several endpoints can be observed: death, transfer to haemodialysis and renal transplantation. In our study, we were interested in modelling the time from the entrance in the peritoneal dialysis program until the occurrence of the event of interest, death, in the presence of competing risks (transfer to haemodialysis and renal transplantation). Data from all patients included in the peritoneal dialysis program (Hospital Geral de Santo Antnio, Centro Hospitalar do Porto, Porto, Portugal) between October 1985 and June 2011 were analyzed. Regression models based on cause-specific hazard and hazard of the subdistribution were performed, considering time-independent (gender, age, diabetes and first treatment) and time-dependent covariates (hospitalizations and peritonitis) and the estimates obtained by such models were examined and discussed.

Keywords: competing risks, cause-specific hazard, hazard of the subdistribution, peritoneal dialysis
AMS subject classifications: 62P10

Bibliography

[1] Andersen, P.K., Geskus, R.B., de Witte, T. and Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. *Int. J. Epidemiol* 41(3), 861–870.
[2] Gooley, T.A., Leisenring, W., Crowley, J. and Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statist. Med.* 18, 695–706.
[3] Pintilie, M. (2007). Analysing and interpreting competing risk data. *Statist. Med.* 26, 1360–1367.
[4] Klein, J.P. (2006). Modelling competing risks in cancer studies. *Statist. Med.* 25, 1015–1034.
[5] Pintilie, M. (2006). *Competing risks. A practical perspective.*, John Wiley and Sons, New Jersey.

Accelerated failure time model for repairable systems

Petr Novák[1]

[1] *Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic*

Abstract

When studying the service record of a device which is a subject to degradation, we want to estimate the time-to-failure distribution for maintenance optimization. The dependency of the failure time distribution on applicable regression variables can be described with a suitable model. For instance, we may use the number of repairs and maintenance actions or their cost as time-varying covariates. For this situation, the Cox proportional hazards model has been suggested, with the repairs and maintenance actions influencing the hazard function multiplicatively. Alternatively, we can use the Accelerated failure time model, where the covariates cause the internal time of the device to flow faster or slower. In this work we describe such models and demonstrate their application on real data.

Keywords: Reliability analysis, Repair models, Regression, Accelerated Failure Time model.
AMS subject classifications: 62N02.
Acknowledgements: This work was supported by the grant SVV 261315/2013.

Robust multivariate process control of multi-way data with root cause analysis

Peter Scheibelhofer[1,2], Günter Hayderer[2] and Ernst Stadlober[1]

[1] *Graz University of Technology, Austria*
[2] *ams AG, Unterpremstätten, Austria*

Abstract

The evaluation of the manufacturing process condition is a crucial challenge in modern semiconductor fabrication. With growing complexity large numbers of process variables are recorded during equipment operations of every process step. For monitoring these processes, traditional fault detection and classification methods were implemented, but they are mostly univariate. Multivariate techniques such as Principal Component Analysis and Hotellings T^2 are capable of advanced process control but are mainly applied on statistically calculated indicators such as means or standard deviations of one wafer over its process time. Thereby, information of the time variation of the variables is omitted. In this work, we present a generalized methodology for multivariate process control that considers the whole recorded information of a wafer by using multi-way principal component analysis (MPCA). The use of Hotellings T^2 statistics makes outcomes easy to monitor as it can be summarized into one control chart. By grouping similar variables into reasonable functional groups and by applying decomposition methods for the T^2 signal, a root cause analysis is possible. Furthermore, special attention is paid on the robustness of the MPCA and T^2 procedure as an analysis independent of frequently observed outliers is crucial. In a case study of production data from Austrian semiconductor manufacturer ams AG an observed production machine error can be detected and its root cause can be tracked down successfully.

Keywords: fault detection, multivariate process control, multi-way principal component analysis, robust statistics
AMS subject classifications: 62P30

Developing statistical methodologies for anthropometry

Guillermo Vinué[1]

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

Abstract

Fitting Ready To Wear clothes is a basic problem for customer and apparel companies. One of the most important problems to develop new patterns and grade to all sizes is the lack of updated anthropometric data. For this reason, national administrations and industrial groups from the clothing sector in different countries have been fostering several national anthropometric surveys in recent years. In 2006 the Spanish Ministry of Health promoted a 3D anthropometric study of the Spanish female population. This survey aimed to generate anthropometric data from Spanish women for the clothing industry [1]. A sample of 10.415 Spanish females from 12 to 70 years old randomly selected from the official Postcode Address File was measured. The obtained anthropometric data constitute valuable information to understand the body shape of the population. A very important challenge is to define an optimal sizing system. A sizing system classifies a specific population into homogeneous subgroups based on some key body dimensions. Our research group has developed some clustering methodologies using some of the ideas of [4, 5], among others. On the other hand, the shape of the 10.415 women is described by using a set of correspondence points. In this case, we have used the statistical shape analysis [3] to divide the population into efficient sizes according to their shape. In the multivariate accommodation problem, a small group of representative cases (human models) which represents the anthropometric variability of the target population is commonly used. The appropriate selection of this small group is critical in order to accommodate a certain percentage of the population. We use the archetypal analysis [2] to that end. The archetypes returned by the archetypal analysis are a convex combination of the sampled individuals, but they are not necessarily observed individuals. However, in human modeling is crucial that the archetypes are real people. An algorithm inspired by the Partitioning Around Medoids (PAM) clustering algorithm to obtain necessarily observed individuals, which we call archetypoids, has been proposed. All the just mentioned statistical methodologies use the anthropometric data of the Spanish survey. They are gathered together in an R package called *Anthropometry*, that will be available soon.

Keywords: Anthropometric data, Clustering, Statistical shape analysis, Archetypal analysis
AMS subject classifications: 62P30.
Acknowledgements: This research has been partially supported by grants TIN2009–14392–C02–01 and TIN2009–14392–C02–02.

Bibliography

[1] Alemany, S., González, J.C., Nácher, B., Soriano, C., Arnáiz, C. and Heras, A. (2010). Anthropometric survey of the Spanish female population aimed at the apparel industry. *Proceedings of the International Conference on 3D Body Scanning Technologies*, Lugano, Switzerland.

[2] Cutler, A. and Breiman, L. (1994). Archetypal Analysis. *Technometrics* 36 (4), 338–347.

[3] Dryden, I.E. and Mardia, K.V. (1998). *Statistical Shape Analysis*, John Wiley & Sons, Chichester.

[4] McCulloch, C., Paal, B. and Ashdown, S. (1998). An optimization approach to apparel sizing. *Journal of the Operational Research Society* 49, 492-499.

[5] van der Laan, M.J. and Pollard, K.S. (2003). A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. *Journal of Statistical Planning and Inference* 117, 275–303.

Hidden Markov models in modeling time series of earthquakes

Katerina Orfanogiannaki[1] and Dimitris Karlis[2]

[1] *Institute of Geodynamics, National Observatory of Athens, Greece*
[2] *Department of Statistics, Athens University of Economics and Business, Greece*

Abstract

Discrete valued Hidden Markov Models (HMMs) are used to model time series of event counts in several scientific fields like genetics, engineering, seismology and finance ([1]). Typically the model consists of two parts: the observed sequence of event counts and an unobserved (hidden) sequence of states that consist a Markov chain. In the univariate case each state is characterized by a different univariate distribution and the progress of the Hidden process from state to state is controlled by a transition probability matrix. Conditional on the state of the model at a specific point of time, the distribution of the observation at that point is fully specified. We have assumed both Poisson (PHMMs) and negative binomial (NBHMMs) distribution to dominate each state. An extension of the univariate HMMs to the multivariate case can be made by assuming a multivariate discrete distribution associated with each state. The known as multivariate Poisson distribution in [2], is incorporated into the model to allow as modeling multivariate discrete valued time series. We examine properties of the model and propose inference. Maximum likelihood estimators of the models' parameters are derived using an EM algorithm. Univariate PHMMs and NBHMMs and a bivariate Poisson HMM are

applied to earthquake data from Sumatra, Indonesia. On 26 December 2004 and 28 March 2005 occurred two of the largest earthquakes of the last 40 years between the Indo-Australian and the southeastern Eurasian plates with moment magnitudes $Mw = 9.1$ and $Mw = 8.6$ respectively. HMMs models are used for identifying temporal patterns in the time series of the two mainshocks. Each time series consists of earthquake counts, in different time units (days, two-day periods, five-day periods), in the regions determined by the aftershock zones of the two mainshocks. In addition, a bivarite Poisson HMM is applied to jointly model the two time series and the correlation between them is estimated.

Keywords: Hidden Markov Models, Poisson, Negative Binomial, Multivariate Poisson, earthquake counts
AMS subject classifications: 62M10
Acknowledgements: This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

Bibliography

[1] MacDonald, I.L., and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete valued Time Series*, Chapman and Hall, London.
[2] Johnson, N., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, Wiley, NY.

On sample selection models and skew distributions

Emmanuel O. Ogundimu[1] **and Jane L. Hutton**[1]

[1]*Department of Statistics, University of Warwick, UK*

Abstract

Scores arising from questionnaires often follow asymmetric distributions, on a fixed range. This can be due to scores clustering at one end of the scale or selective reporting. Sometimes, the scores are further subjected to sample selection resulting in partial observability. Thus, methods based on complete cases for skew data are inadequate for the analysis of such data and a general sample selection model is required. Heckman proposed a full maximum likelihood estimation method under the normality assumption for sample selection problems, and parametric and non-parametric extensions have been proposed. A general selection distribution for a vector $\mathbf{Y} \in \mathbb{R}^p$ has a PDF f_Y given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}^\star}(\mathbf{y}) \frac{P(\mathbf{S}^\star \in \mathbf{C} \,|\, \mathbf{Y}^\star = y)}{P(\mathbf{S}^\star \in \mathbf{C})},$$

where $\mathbf{S}^\star \in \mathbb{R}^q$ and $\mathbf{Y}^\star \in \mathbb{R}^p$ are two random vectors, and $\mathbf{C}$ is a measurable subset of $\mathbb{R}^q$. We use this generalization to develop a sample selection model with underlying skew-normal distribution. A link is established between the continuous component of our model log-likelihood function and an extended version of a generalized skew-normal distribution. This link is used to derive the expected value of the model, which extends Heckman's two-stage method. Finite sample performance of the maximum likelihood estimator of the model is studied via Monte Carlo simulation. The model parameters are more precisely estimated under the new model, even in the presence of moderate to extreme skewness, than the Heckman selection models. Application to data from a study of neck injuries, where the responses are substantially skew, successfully discriminates between selection and inherent skewness. We also discuss computational and identification issues, and provide an extension of the model using underlying skew-t distribution.

Keywords: Generalized Sample selection, Missing data, Closed Skew-normal distribution, Closed Skew-t distribution.
AMS subject classifications: 62D99

Bibliography

[1] Heckman, J. (1976) . The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–492.

[2] Copas, J. B. and Li, H. (1997). Inference for non-random samples J. R. Statist. Soc. B 59, 55–95.

BCa-JaB method as a diagnostic tool for linear regression models

Ufuk Beyaztas[1] **and Aylin Alin**[1]

[1] *Department of Statistics, Dokuz Eylul University*

Abstract

Jackknife-after-bootstrap (JaB) has first been proposed by [3] then used by [4] and [1] to detect influential observations in linear regression models. This method uses the percentile confidence interval to provide cut-off values for the measures. In order to improve JaB, we propose using Bias Corrected and accelerated (BCa) confidence interval introduced by [2]. In this study, the performance of BCa-JaB and conventional JaB methods are compared for DFFITS, Welsch's distance, modified Cook's distance and t-star statistics. Comparisons are based on both real world examples and simulation study. The results reveal that under considered scenarios proposed method provides more symmetric threshold values which give more accurate and reliable results.

Keywords: Bootstrap, BCa confidence interval, influential observation, regression diagnostics.
AMS subject classifications: 62F40; 62G09; 62J05; 62J20

Bibliography

[1] Beyaztas, U. and Alin, A. (2013). Jackknife-after-bootstrap method for detection of influential observations in linear regression models. *Communications in Statistics: Simulation and Computation* 42, 1256–1267.

[2] Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the Statistical Association* 82, 171–185.

[3] Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions. *Journal of the Royal statistical Society* 54, 83–127.

[4] Martin, M.A. and Roberts, S. (2010). Jackknife-after-bootstrap regression influence diagnostics. *Journal of Nonparametric Statistics* 22, 257–269.

The zero area Brownian bridge

Maik Görgens[1]

[1]*Department of Mathematics, Uppsala University, P.O.Box 480, 751 06 Uppsala, Sweden*

Abstract

We consider the Brownian motion W on the interval $[0,1]$. The Brownian bridge B arises from the Brownian motion by pinning W_1 down to 0, i.e., the Brownian bridge arises by conditioning the Brownian motion to fulfill $W_1 = 0$. We condition the Brownian bridge further by requiring $\int_0^1 B_s ds = 0$. We call the resulting Gaussian process on $[0,1]$ zero area Brownian bridge and denote it by M. We will study properties of M and give anticipative as well as non-anticipative representations. Our main tool to access the zero area Brownian bridge is to consider the associated operator $u : L_2([0,1]) \to C([0,1])$ of the Brownian motion, where u is given by

$$(uf)(s) = \int_0^s f(x)dx, \qquad f \in L_2([0,1]).$$

Then (a version of) the Brownian motion is given by

$$W_s = \sum_{i=0}^{\infty} \xi_i (ue_i)(s),$$

where $(e_i)_i$ is an orthonormal basis in $L_2([0,1])$ and $(\xi_i)_i$ is a series of independent standard normal random variables (cf. [2]). The zero area Brownian bridge is the Gaussian process associated to the operator $v : H \to C([0,1])$, where v is the restriction of u to a suitable subspace $H \subset L_2([0,1])$.

Keywords: Gaussian processes, Conditioning, Brownian bridge, Series expansions
AMS subject classifications: 60G15, 60H10, 60J65

Bibliography

[1] Görgens, M. (2013) Conditioning of Gaussian processes and a zero area Brownian bridge. Preprint, arXiv:1302.4186.

[2] Mörters, P. and Peres, Y. (2010). *Brownian motion*, Cambridge University Press.

Intervention in Ornstein-Uhlenbeck SDEs

Alexander Sokol[1]

[1]*University of Copenhagen, Denmark*

Abstract

We introduce a notion of intervention for stochastic differential equations and a corresponding causal interpretation. For the case of the Ornstein-Uhlenbeck SDE, we show that the SDE resulting from a simple type of intervention again is an Ornstein-Uhlenbeck SDE. We discuss criteria for the existence of a stationary distribution for the solution to the intervened SDE. We illustrate the effect of interventions by calculating the mean and variance in the stationary distribution of an intervened process in a particularly simple case.

Keywords: Causality, Intervention, SDE, Ornstein-Uhlenbeck process, Stationary distribution.
AMS subject classifications: 60G15.
Acknowledgements: The development of the notion of intervention for SDEs is joint work with my thesis advisor, Niels Richard Hansen, whom I also thank for valuable discussions and advices.

Bibliography

[1] Aalen, Odd O., Røysland, Kjetil; Gran, Jon Michael. Causality, mediation and time: A dynamic viewpoint. *J. R. Statist. Soc. A* 175 (2012), no. 4, 831–861.
[2] Cross, G. W. Three types of matrix stability. *Linear Algebra and Appl.* 20 (1978), no. 3, 253–263.
[3] Florens, Jean-Pierre; Fougere, Denis. Noncausality in continuous time. *Econometrica* 64 (1996), no. 5, 1195–1212.
[4] Gégout-Petit, Anne; Commenges, Daniel. A general definition of influence between stochastic processes. *Lifetime Data Anal.* 16 (2010), no. 1, 33–44.
[5] Hershkowitz, Daniel; Keller, Nathan. Positivity of principal minors, sign symmetry and stability. *Linear Algebra Appl.* 364 (2003), 105–124.
[6] Horn, Roger A.; Johnson, Charles R. Matrix analysis. *Cambridge University Press, Cambridge*, 1985.
[7] M. Jacobsen: A Brief Account of the Theory of Homogeneous Gaussian Diffusions in Finite Dimensions, in "Frontiers in Pure and Applied Probability 1", p. 86-94, 1993.
[8] Pearl, Judea. Causality. Models, reasoning, and inference. Second edition. *Cambridge University Press, Cambridge*, 2009.
[9] Petrović, Ljiljana; Dimitrijević, Sladjana. Invariance of statistical causality under convergence. *Statist. Probab. Lett.* 81 (2011), no. 9, 1445–1448.
[10] Protter, Philip E. Stochastic integration and differential equations. Second edition. Version 2.1. Corrected third printing. *Springer-Verlag, Berlin*, 2005.

[11] Rogers, L. C. G.; Williams, David. Diffusions, Markov processes, and martingales. Vol. 1. Foundations. *Cambridge University Press*, Cambridge, 2000.

[12] Zakai, Moshe; Snyders, Jakov. Stationary probability measures for linear differential equations driven by white noise. *J. Differential Equations* 8 (1970), 27–33.

Some properties of a class of continuous time moving average processes

Andreas Basse-O'Connor[1]

[1]*Department of Mathematics, Aarhus University, Denmark.*
E-mail: basse@imf.au.dk

Abstract

In discrete time, moving average processes play an important role in time series analysis. A moving average is a process $\{X_n\}_{n\in\mathbb{N}}$ of the form $X_n = \sum_{k=-\infty}^{n} \phi_{n-k} Z_k$ where $\{\phi_k\}_{k\in\mathbb{N}}$ is a deterministic sequence of real numbers and $\{Z_k\}_{k\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables. In continuous time, moving averages are processes $X = \{X_t : t \in \mathbb{R}_+\}$ of the form

$$X_t = \int_{-\infty}^{t} \phi(t-s)\, dZ_s \tag{2}$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}$ is a deterministic function and $Z = \{Z_t : t \in \mathbb{R}\}$ is a process with stationary and independent increments (a so-called Lévy process). In this work we will consider a continuous time moving average X of the form (2) in the the case where the kernel function ϕ is the gamma density, i.e. $\phi(t) = e^{-\lambda t} t^{\gamma-1}$. We will derive necessary and sufficient conditions for X to be well-defined, that is, for the existence of the stochastic integrals (2). In some cases X has very irregular sample paths, e.g. they are unbounded on every bounded interval. We give necessary and sufficient conditions for X to have the following type of regularity: almost all sample paths are of bounded variation, or more generally, the process is a semimartingale. These two condition corresponds to that stochastic integrals of the form $\int_0^t Y_s\, dX_s$ are well-defined in the Lebesgue–Stieltjes sense or in the Itô sense, respectively. Our work uses the recent results [2, 3, 4]. Finally let us mention that the gamma kernel has been extensively used to build stochastic models for turbulence; see [1] and the reference therein.

Keywords: Moving averages, gamma density, bounded variation, semimartingales
AMS subject classifications: 60G48; 60H05; 60G51; 60GH17

Bibliography

[1] Barndorff-Nielsen, O. E. (2012). Notes on the gamma kernel. *Thiele Research Reports;* No. 03. Available at http://math.au.dk/publs?publid=946.

[2] Basse, A. and Pedersen, J. (2009). Lévy driven moving averages and semimartingales. *Stochastic Process. Appl.* 119(9), 2970-2991.

[3] Basse-O'Connor, A. and Rosiński, J. (2012). Structure of infinitely divisible semimartingales. arXiv:1209.1644v2 [math.PR].

[4] Basse-O'Connor, A. and Rosiński, J. (2013). Characterization of the finite variation property for a class of stationary increment infinitely divisible processes. *Stochastic Process. Appl.* doi:10.1016/j.spa.2013.01.014.

Gibbs point process approximation based on Stein's method

Dominic Schuhmacher[1] and Kaspar Stucki[1]

[1] *Institute of Mathematical Statistics and Sctuarial Science, University of Bern, Switzerland*

Abstract

Gibbs point processes are widely used both in spatial statistics and in statistical mechanics, as they allow a very flexible modelling of interaction between the points. However, a notorious difficulty with Gibbs processes is that in most cases of interest their densities can only be specified up to normalizing constants, which typically renders explicit calculations, e.g. of the total variation distance between two such processes, difficult.

Based on the results of Schuhmacher and Stucki [1] we obtain upper bounds for the total variation distance between the distributions of two Gibbs point processes in a very general setting. Applications are provided to various well-known processes and settings from spatial statistics and statistical physics, including the comparison of two Lennard-Jones processes or hard core approximation of an area interaction process.

The proof of the main results is based on Stein's method. We construct an explicit coupling between two spatial birth-death processes to obtain Stein factors, and employ the Georgii-Nguyen-Zessin equation for the total bound.

Keywords: Conditional intensity, pairwise interaction process, birth-death process, Stein's method, total variation distance.

AMS subject classifications: Primary 60G55; Secondary 60J75, 82B21.

Bibliography

[1] Dominic Schuhmacher and Kaspar Stucki. On bounds for Gibbs point process approximation. *Preprint*, 2012. Available at http://arxiv.org/abs/1207.3096.

Optimal designs for discriminating between functional linear models

Verity Fisher[1] **and Dave Woods**[1]

[1] *University of Southampton, UK*

Abstract

Improvements in online measuring and monitoring have facilitated an increase in the number of observations that can be taken on each experimental unit in industrial and scientific experiments. Examples include biometry, chemistry, psychology and climatology. It can often be assumed that the data for each run are generated by a smooth underlying function. We are interested in how changes to the levels of the controllable factors influence these functions. Often a semi-parametric model is assumed for the response, with relatively simple polynomial models describing the treatment effects.
Methods are presented for the design of experiments with functional data when the aim is to discriminate between linear models for the treatment effect. We develop an extension of the T-optimality criterion to functional data for discriminating between two competing models. The methodology is motivated by an example from Tribology and assessed via simulation studies to calculate the sensitivity and specificity of the resulting analyses.

Keywords: Optimal design, model discrimination, functional data, prediction
AMS subject classifications: 62K05
Acknowledgements: This work was supported by the UK Engineering and Physical Sciences Research Council and GlaxoSmithKline, the former through a PhD studentship (Fisher) and a Research Fellowship (Woods), the latter through a Research Fellowship (Fisher).

A stochastic optimization method for constructing optimal block designs with linear constraints

Alena Bachratá[1] **and Radoslav Harman**[1]

[1]*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia*

Abstract

We propose a stochastic optimization method related to simulated annealing for constructing efficient designs of experiments under a broad class of linear constraints on the design weights. The linear constraints can represent restrictions on various types of "limits" associated with the experiment.
To illustrate the method we computed D-, A-, and E-optimal designs for estimating a set of treatment contrasts in the case of block size-two experiments with upper limits on the number of replications of each non-control treatment.

Keywords: stochastic optimization, design of experiments, linear constraints, block designs
AMS subject classifications: 62K05, 62K10

Directed random graphs and convergence to the Tracy-Widom distribution

Takis Konstantopoulos[1] **and Katja Trinajstić**[1]

[1] *Department of Mathematics, Uppsala University, Sweden*

Abstract

We consider a directed random graph on the 2-dimensional integer lattice, placing independently, with probability p, a directed edge between any pair of distinct vertices (i_1, i_2) and (j_1, j_2), such that $i_1 \leq j_1$ and $i_2 \leq j_2$. Let $L_{n,m}$ denote the maximum length of all paths contained in an $n \times m$ rectangle. The asymptotic distribution for a centered/scaled version of $L_{n,m}$, for fixed m, as $n \to \infty$, was derived in [2]. Here, we address the problem of finding the limit when both n and m tend to infinity, so that $m \sim n^a$. We make a sequence of transformations in order to exhibit a resemblance of our model to a last passage percolation model. This requires the use of suitably defined regenerative points (called skeleton points), together with a number of pathwise and probabilistic bounds. Making use of a Komlós-Major-Tusnády coupling, as in [1], with a last-passage Brownian percolation model, we are able to prove that, for $a < 3/14$, the asymptotic distribution is the Tracy-Widom distribution.

Keywords: Random graph, Last passage percolation, Strong approximation, Tracy-Widom distribution
AMS subject classifications: 05C80, 60F17, 60K35, 06A06

Bibliography

[1] Bodineau, T. and Martin, J. (2005). A universality property for last-passage percolation paths close to the axis. *Electron. Commun. Probab.* 10, 105–112.

[2] Denisov, D., Foss, S. and Konstantopoulos, T. (2012). Limit theorems for a random directed slab graph. *Ann. Appl. Probab.* 22, 702–733.

Some remarks on normal conditionals and normal projections

Barry C. Arnold[1] **and B.G. Manjunath**[2]

[1] *University of California Riverside, USA*
[2] *CEAUL and DEIO, FCUL, University of Lisbon, Portugal*

Abstract

It is always possible to construct a d-dimensional non-normal distribution having any finite number of normal projections and all $(d-1)$ dimensional marginals normal. Also, there can exist d-dimensional non-normal distribution with all conditional distributions being normal. In the present note we introduce two new characterization of the classical d-dimensional normal distribution. (1) Having normal conditionals and a finite number of normal projections uniquely characterizes the classical d-dimensional normal distribution. (2) Having normal conditionals and each of $(d-1)$ coordinate random variables having a one dimensional normal distribution is sufficient to ensure that the d-dimensional distribution has to be classical normal.

Keywords: non-normal distributions, normal conditionals, normal marginals, linear transformation
AMS subject classifications: 62E10 and 62E15
Acknowledgements:[2]Research enabled through a grant from the Foundation for Science and Technology (FCT), project SFRH/BPD/72184/2010 and FCT / PTDC /MAT/101736/2008 "EXTREMA: EXTREMES IN TODAY'S WORLD".

Bibliography

[1] Barry C. Arnold, Enrique Castillo and Jose M. Sarabia, (1999). *Conditional Specification of Statistical Models*, Springer, New York.
[2] Hamedani, G.G. and Tata, M.N. (1975). On the determination of the bivariate normal distribution from distributions of linear combinations of the variables. *The American Mathematical Monthly* 82, 913-915.
[3] Manjunath, B.G. and Parthasarathy, K.R. (2012). A note on Gaussian distributions in R^n. *Proc. Indian Acad. Sci. (Math. Sci.)* 122 (4), 635–644.
[4] Stoyanov, J. (1997). *Counterexamples in Probability (2nd ed.)*, Wiley, New York.

Constructing hierarchical copulas using the Kendall distribution function

Eike Christian Brechmann[1]

[1]*Center for Mathematical Sciences, Technische Universität München*

Abstract

While there is substantial need for dependence models in higher dimensions, most existing models are rather restrictive and barely balance parsimony and flexibility. In this talk, the class of hierarchical Kendall copulas is proposed as a new approach to tackle these problems. By aggregating dependence information of non-overlapping groups of variables in different hierarchical levels using the Kendall distribution function, hierarchical Kendall copulas provide a new and attractive option to model dependence patterns between large numbers of variables
In particular, let $U_1, ..., U_n$ be uniform random variables, C_0 a d-dimensional copula and $C_1, ..., C_d$ copulas of dimension $n_1, ..., n_d$, where $n = \sum_{i=1}^{d} n_i$. Further, define $m_i = \sum_{j=1}^{i} n_j$ for $i = 1, ..., d$ and $m_0 = 0$. Then the random vector $(U_1, ..., U_n)'$ is said to be distributed according to the two-level hierarchical Kendall copula $C_{\mathcal{K}}$ with nesting copula C_0 and cluster copulas $C_1, ..., C_d$ if

1. $(U_{m_{i-1}+1}, ..., U_{m_i})' \sim C_i$ for all $i = 1, ..., d$, and
2. $(V_1, ..., V_d)' \sim C_0$, where $V_i := K_i(C_i(U_{m_{i-1}+1}, ..., U_{m_i})) \sim U(0,1)$ for all $i = 1, ..., d$ and K_i denotes the Kendall distribution function corresponding to C_i, that is the multivariate probability integral transform

$$K_i(t) := P(C_i(U_{m_{i-1}+1}, ..., U_{m_i}) \leq t) \quad \text{for } t \in [0,1].$$

This mimics the standard copula approach with univariate margins. The cluster copulas $C_1, ..., C_d$ and the nesting copula C_0 can be chosen independently from any class of copulas. Furthermore, the definition can also easily be extended to an arbitrary number of levels.
The talk explicitly discusses properties as well as inference techniques for hierarchical Kendall copulas, in particular, simulation, estimation and model selection. A closed-form sampling algorithm is derived for Archimedean, Plackett and Archimax copulas, while for general copulas an approximative method is proposed. For estimation, a sequential and a joint approach are discussed and compared in an extensive simulation study.
In an application, a three-level hierarchical Kendall copula will be used to conduct a systemic risk stress testing exercise of the international financial sector, providing a ranking of systemically important institutions.

Keywords: multivariate copula, hierarchical copula, Kendall distribution function
AMS subject classifications: 62H20

Bibliography

[1] Brechmann, E. C. (2012). Hierarchical Kendall copulas: Properties and inference. Preprint, http://arxiv.org/abs/1202.1998.

Minimum description length principle and distribution complexity of spherical distributions

Bono Nonchev[1]

[1]*Faculty of Mathematics and Informatics, University of Sofia, Bulgaria*

Abstract

The application of the MDL principle to discern from which distribution a sample originates is discussed with the focus is on the general class of spherical distributions. Their trivial generalization the elliptical distributions are widely used in financial theory and have properties that enable us to calculate a closed form solution of their distribution complexity.

The MDL principle and its codelength/model interpretation is discussed first, as well as its application in model selection. Then the NML model is introduced as a suitable choice and its equivalent formulation as the model complexity is explored. After that the distribution complexity is presented as a solution of the problem of infinite model complexity, with the rest of the paper exposing the main result - the calculation of the distribution complexity for spherical distributions.

The analytical formulas for the distribution complexity are explicitly shown in three cases - the Gaussian distribution, the Student-T distribution and the Laplace distribution. Thoughts on their interpretation of the change of complexity with the size of the sample are presented with a somewhat surprising characterization of the NML model for the spherical distributions that has potential impact on robust estimation.

Keywords: MDL, Model Selection, Complexity, Distribution Selection, Spherical distributions, Student-T distribution, Laplace distribution

AMS subject classifications: 94A17, 62B10, 62F03

Acknowledgements: This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014).

Bibliography

[1] Peter Grünwald, Jay Injae Myung, and Mark Pitt. *Advances in Minimum Description Length: Theory and Applications.* The MIT Press, April 2005.

[2] Robert Kass and Adrian Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.
[3] Andrey Nikolaevich Kolmogorov. On Tables of Random Numbers. *Sankhya: The Indian Journal of Statistics, Series A*, 25(4):369–376, December 1963.
[4] Guoqui Qian and Hans Künsch. On Model Selection in Robust Linear Regression. 1996.
[5] Jorma Rissanen. MDL Denoising. *IEEE Transactions on Information Theory*, 46(7):2537–2543, 2000.
[6] Jorma Rissanen. *Information and Complexity in Statistical Modeling (Information Science and Statistics)*. Springer, January 2007. ISBN 0387366105.
[7] Yuri Shtarkov. Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, (3):175–186, July .

Skewed sub-Gaussian multivariate distribution

Teodosi Geninski[1], **Ivan Mitov**[2], **Zari Rachev**[3]

[1]*Faculty of Mathematics and Informatics, Sofia University, Bulgaria*
[2]*FinAnalytica Inc.*
[3]*Stony Brook University*

Abstract

Normal variance mixture models are used as an extension of the Gaussian framework to allow heavier tails and add flexibility to the Wiener processes time concept. The Sub-Gaussian model is a typical representative of this class. It is a parametric sub-class of the multivariate α-stable distribution which is an elliptical, infinitely divisible and has a tractable representation of its characteristic function. It possesses heavy tails but it is also a symmetric distribution.
To overcome the latter drawback a ρ-weighted, univariate, α-stable skewness component is introduced. The domain of ρ and its connection to the skewness and the dependence structure are explored as well as some of the border cases. By varying ρ from 0 to 1 the distribution transforms from a regular Sub-Gaussian to multivariate α-stable with independent and not necessary symmetric components.
Application to a real-world financial assets data is provided together with fitting techniques and comparison of the Sub-Gaussian and the Skewed Sub-Gaussian distribution.

Keywords: Variance mixture, Multivariate stable models, Sub-Gaussian model, Asymmetric distributions
AMS subject classifications: 60E07, 62P05, 62E17

Acknowledgements: This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014).

Bibliography

[1] Rachev S. and Mittnik S., (2000), Stable Paretian Models in Finance, Wiley
[2] Samorodnitsky G., Taqqu M. S., (1994), Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance, Chapman & Hall.

Author index

Affiliation and Contacts

KEYNOTE SPEAKER	AFFILIATION	E-MAIL ADDRESS
Bojan Basrak	Department of Mathematics, University of Zagreb, Croatia	bbasrak@math.hr
Nikolai N. Leonenko	School of Mathematics, Cardiff University, United Kingdom	LeonenkoN@cardiff.ac.uk
Jürgen Pilz	Alpen-Adria Universitaet Klagenfurt, Austria	Juergen.Pilz@uni-klu.ac.at
Johan Segers	Université catholique de Louvain, Belgium	johan.segers@uclouvain.be
Michael Sørensen	Department of Mathematical Sciences, University of Copenhagen, Denmark	michael@math.ku.dk

COUNTRY	PARTICIPANT	AFFILIATION	E-MAIL ADDRESS
Austria	Peter Scheibelhofer	Graz University of Technology (PhD student) and ams AG	peter.scheibelhofer@ams.com
	Lukas Steinberger	Department of Statistics and OR, University of Vienna	lukas.steinberger@univie.ac.at
Belgium	Joke Durnez	Ghent University	joke.durnez@ugent.be
	Rudolf Schenk	Université catholique de Louvain	rudolf.schenk@uclouvain.be
Bulgaria	Bono Nonchev	Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski"	bono@nonchev.info
	Teodosi Geninski	Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski"	teodosi.g@gmail.com
Croatia	Danijel Grahovac	Department of Mathematics, J.J. Strossmayer University of Osijek	dgrahova@mathos.hr
	Snježana Lubura	Department of Mathematics, University of Zagreb	snjezana.lubura@math.hr
Czech Republic	Petr Novak	Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague	novakp@karlin.mff.cuni.cz

COUNTRY	PARTICIPANT	AFFILIATION	E-MAIL ADDRESS
Czech Republic	Katarína Starinská	Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague	starinskak@gmail.com
Denmark	Andreas Basse-O'Connor	Department of Mathematics, University of Aarhus	basse@imf.au.dk
	Alexander Sokol	Department of Mathematical Sciences, University of Copenhagen	alexander@math.ku.dk
Finland	Leena Annukka Pasanen	Department of Mathematical Sciences, University of Oulu	Leena.Pasanen@oulu.fi
	Lauri Viitasaari	Aalto University School of Science	lauri.viitasaari@aalto.fi
France	Sarah Ouadah	Laboratoire de Statistique Théorique et Appliquée (LSTA), University Paris 6	sarah.ouadah@upmc.fr
	Eric Sibony	Tlcom ParisTech	esibony@gmail.com
Germany	Eike Christian Brechmann	Technische Universität München, Zentrum Mathematik, Lehrstuhl fr Mathematische Statistik	brechmann@ma.tum.de
	Philip Preuß	Ruhr-Universität Bochum, Fakultät für Mathematik, Lehrstuhl fr Stochastik	philip.preuss@ruhr-uni-bochum.de
Greece	Katerina Orfanogiannaki	Department of Statistics, Athens University of Economics and Business	korfanogiannaki@gmail.com
	Panagiotis Papastamoulis	Department of Statistics & Insurance Science, University of Pireaus	papapast@yahoo.gr
Hungary	Kinga Sikolya	Faculty of Informatics, University of Debrecen, Bolyai Institute, University of Szeged	sikolya.kinga@inf.unideb.hu
	Ferenc Rrosi	University of Szeged	rarosiferenc@gmail.com
Italy	Manuela Cattelan	Department of Statistical Sciences, University of Padova	manuela.cattelan@stat.unipd.it
	Francesco Stingo	Department of Biostatistics, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA	fra.stingo@gmail.com
The Netherlands	Juan Juan Cai	Dept. of Applied Mathematics, Delft University of Technology	caicaij@gmail.com
	Botond Szabo	Eindhoven University of Technology	b.szabo@tue.nl
Poland	Alicja Szabelska	Department of Mathematical and Statistical Methods, Poznan University of Life Sciences	alicja.szabelska@fgcz.uzh.ch

COUNTRY	PARTICIPANT	AFFILIATION	E-MAIL ADDRESS
Poland	Piotr Szulc	Institute of Mathematics and Computer Science, Wroclaw University of Technology	piotr.a.szulc@pwr.wroc.pl
Portugal	B.G. Manjunath	Department of Statistics and Applications, Faculty of Sciences, Lisbon University	bgmanjunath@gmail.com
	Laetitia Da Costa Teixeira	Faculty of Sciences and Institute of Biomedical Sciences Abel Salazar, University of Porto	laetitiateixeir@gmail.com
Romania	Mircea Dumitru	Faculty of Mathematics and Informatics, University Bucharest, SUPELEC - Univ. Paris Sud	Mircea.Dumitru@lss.supelec.fr
	Mircea Dragulin	Faculty of Mathematics and Informatics, University Bucharest	mircea.mate@yahoo.com
Russia	Abakirova Aygul	Moscow State University	abakirova@gmail.com
	Ekaterina Krymova	Institute for Information Transmission Problems, Moscow	ekkrym@gmail.com.
Slovakia	Gábor Szücs	Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics, Comenius University	szucs@fmph.uniba.sk
	Alena Bachratá	Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics, Comenius University	alena.bachrata@fmph.uniba.sk
Slovenia	Žiga Kotnik	Faculty of Administration, University of Ljubljana	ziga.kotnik@fu.uni-lj.si
Spain	Nuria Torrado-Robles	Department of Statistical Methods, University of Zaragoza	nuria.torrado@gmail.com
	Guillermo Vinué	Department of Statistics and Operational Research, Faculty of Mathematics, Universidad de Valencia	Guillermo.Vinue@uv.es
Sweden	Katja Trinajstić	Department of Mathematics, Uppsala University	katja@math.uu.se
	Maik Görgens	Department of Mathematics, Uppsala University	maik@math.uu.se
Switzerland	Kaspar Stucki	Institut for Mathematical Statistics and Actuarial Science, University of Bern	kaspar.stucki@stat.unibe.ch
	Petro Kolesnyk	Institut for Mathematical Statistics and Actuarial Science, University of Bern	petro.kolesnyk@stat.unibe.ch

COUNTRY	PARTICIPANT	AFFILIATION	E-MAIL ADDRESS
Turkey	Ufuk Beyaztas	Department of Statistics, Faculty of Science, Dokuz Eylul University	ufuk.beyaztas@deu.edu.tr
	Çiğdem Güngör	Department of Statistics, Faculty of Natural Sciences, Istanbul Medeniyet University	cgungor@medeniyet.edu.tr
Ukraine	Kostiantyn Ralchenko	Dept. of Probability Theory, Statistics and Actuarial Mathematics, Mechanics and Mathematics Faculty, Taras Shevchenko National University of Kyiv	k.ralchenko@gmail.com
	Alexey Doronin	Dept. of Probability Theory, Statistics and Actuarial Mathematics, Mechanics and Mathematics Faculty, Taras Shevchenko National University of Kyiv	al_doronin@ukr.net
United Kingdom	Emmanuel Ogundimu	Department of Statistics, University of Warwick	O.E.Ogundimu@warwick.ac.uk
	Verity Fisher	Mathematics, University of Southampton	v.fisher@southampton.ac.uk

Sponsors

Bernoulli Society for Mathematical Statistics and Probability

Ministry of Science, Education and Sports of the Republic of Croatia

Croatian Academy of Sciences and Arts

Croatian Chamber of Economy

Osijek-Baranja County

City of Osijek

AZ Pension Fund

Konzum

Perutnina

Mehanotehna d.o.o.

Tourist Board, City of Osijek