The 27th Nordic Conference
in Mathematical Statistics

# Abstracts

June 26 - June 29, 2018
Tartu, Estonia

Tartu
2018

# Spatial growth and competition

**Daniel Ahlberg**

*Stockholm University, Sweden, Daniel.ahlberg@math.su.se*

In 1998, Häggström and Pemantle introduced a model for competing growth on $\mathbb{Z}^2$. Motivation for the model was its connection to geodesics in first-passage percolation. They showed that two competing entities of equal strength with positive probability will coexist, in the sense that they may grow to occupy infinitely many sites each. Moreover, they conjectured that coexistence is not possible for competitors of unequal strength. Today, twenty years later, the conjecture remains open, although partial results have been obtained over the years. In this talk we shall review some of these results, including a half-plane version of the conjecture recently settled together with Mia Deijfen and Chris Hoffman.

# Practical problems with tests of cointegration rank with strong persistence and heavy-tailed errors

**Niklas Ahlgren**[1] **and Paul Catani**[2]

[1] *Hanken School of Economics, Finland, niklas.ahlgren@hanken.fi*
[2] *Hanken School of Economics, Finland, paul.catani@hanken.fi*

**Keywords**: cointegration, conditional heteroskedasticity, heavy tails, Monte Carlo simulation, wild bootstrap.

Financial time series have several distinguishing features (referred to as stylised facts). These properties, which include strong persistence, very high persistence in volatility and tail heaviness, are of concern in tests of cointegration.

We consider the $p$-dimensional heteroskedastic cointegrated vector autoregressive (VAR) model of [1]:

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \alpha\rho' D_t + \phi d_t + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $\alpha$ and $\beta$ are $(p \times r)$ matrices of rank $r < p$. The power of asymptotic, bootstrap and wild bootstrap (WB) tests of cointegration rank with strong persistence, very high persistence in volatility and heavy-tailed errors is investigated. We show that tests for cointegration have low power under such conditions. Obtaining high power requires more than 1000 observations, or more than four years of daily observations. The asymptotic and bootstrap tests are unreliable if the errors are heavy-tailed with infinite fourth moment. Monte Carlo simulations indicate that the WB test may be justified with heavy-tailed errors which do not have finite fourth moment.

The tests are applied to daily observations from 2010 to 2016 on the credit default swap (CDS) price and bond spread of US and European investment-grade firms. The empirical results support strong pesistence and very high persistence in volatility. Hill estimates ([2]) for the standardised residuals from the VAR models indicate that the distribution of the errors has heavy tails with finite variance but infinite fourth moment. The WB test accepts cointegration for most firms in the full sample period ($T = 1739$). The evidence for cointegration is weak in sub-sample periods from 2010 to 2013 ($T = 869$) and 2013 to 2016 ($T = 870$).

## References

[1] Cavaliere, G., Rahbek, A., Taylor, A. M. R. (2014). Bootstrap determination of the co-integration rank in heteroskedastic VAR models. *Econometric Reviews* **33**, 606–650.

[2] Hill, J. B. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory* **26**, 1398–1436.

# Ruin Probability for Phase-Type Distributed Claims

**Mohammad Jamsher Ali, Kalev Pärna**

*University of Tartu, Estonia*

In this article, we extend the concept of classical ruin probability of exponentially distributed claims to the hyper-exponentially, i.e. the mixture of exponential distributions, distributed claims. For this, we consider phase-type distribution which is the absorption time distribution in a Markov process where the states are transient except, one is absorbing. We obtain explicit formulas for the ruin probability for hyper-exponentially distributed claims that are mixtures of several exponential distributions. Both, cases with equal and different intensities and mixture weights are considered. Secondly, we cover some issues related to aggregation of claims with purpose to decrease the ruin probability.

## References

[1] S. Asmussen and H. Albrecher, 2010, *Ruin Probabilities*, World Scientific.

[2] R. A. Beezer, 2006, *A First Course in Linear Algebra*.

# On a parametrical estimation for a convolution of exponential densities

**Alexander Andronov**[1], **Nadezda Spiridovska**[2] **and Diana Santalova**[3]

[1] *Saint-Petersburg State University of Civil Aviation, Russia*
[1] *Transport and Telecommunication Institute, Latvia, lora@mailbox.riga.lv*
[2] *Transport and Telecommunication Institute, Latvia, Spiridovska.N@tsi.lv*
[3] *Department of Teacher Education, NTNU, Norway, dianasa@stud.ntnu.no*

**Keywords**: Markov chain, exponential distribution, convolution, Laplace transform.

Broad application of the continuous time Markov chain is caused by exponential distribution properties [2, 3]. The usage of non-exponential distributions lead to considerable analytical difficulties. It is temptingly a using an approximation of arbitrary nonnegative density by a convolution of exponential densities. Two aspects of the problem are considered. Firstly, an approximation of fixed non-negative density. Secondly, the parametrical estimation of the convolution on the basis of given statistical data. Different approaches to such approximation and estimation are considered: the method of the moments, maximum likelihood method, using of Laplace transform of the density. The latter is the least known approach and it proved one's worth as applicable method in the paper [1]. An empirical analysis of different approach has been performed using the simulation. The efficiency of the considered approach is illustrated by the task of the queuing theory.

## References

[1] Andronov, A.M. (2017). On reward rate estimation for a continue time ergodic finite Markov chain. *Advances in Statistics and Simulation* **11(3)**, 407–417.

[2] Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Chapman & Hall, London.

[3] Pacheco, A., Tang, L.C., Prabhu, N.U. (2009). *Markov-Modulated Processes & Semiregenerative Phenomena*. World Scientific, New Jersey - London.

# Stepwise, minimax, and weighted methods for sequential tests of multiple hypotheses

**Michael Baron**[1]

[1] *American University, U.S.A., baron@american.edu*

**Keywords**: Error spending; Familywise error rates; Minimax principle; Sequential probability ratio test; Stepwise testing; Stopping boundaries.

The problem of multiple hypotheses testing often arises in sequential experiments such as sequential clinical trials with multiple endpoints, multichannel change-point detection, acceptance sampling with different criteria of acceptance, etc. In such studies, it is necessary to reach a statistical decision for each individual hypothesis instead of combining them and giving one answer to the resulting composite hypothesis.

Non-sequential methods of multiple comparisons are already well developed. There are Holm, Hommel, Benjamini-Hochberg, Guo-Sarkar, and other methods that can control the familywise error rate or the false discovery rate.

We construct analogues of these methods for sequential data. Combining classical ideas of sequential testing with non-sequential stepwise methods for multiple comparisons, we develop procedures for conducting tests of multiple hypotheses sequentially. Proposed methods control both Type I and Type II familywise error rates, or familywise power, in the strong sense. We attempt to minimize the expected sample size (expected cost) of the experiment under these constraints.

Simultaneous testing of multiple hypotheses can be further optimized by special weights and unequal marginal $\alpha$- and $\beta$-levels that account for the relative difficulty of tested hypotheses. The resulting solution follows the minimax principle that appears naturally for this problem.

Applications of the developed methodology to other optimization problems in multiple comparisons are discussed.

This research is supported by the U.S. National Science Foundation.

## References

[1] Baron, M., MacMillan, L. (2018). Weighted statistics and minimax error spending for testing multiple hypotheses. *Under review.*

[2] De, S. and M. Baron (2012a). Sequential Bonferroni methods for multiple hypothesis testing with strong control of familywise error rates I and II. *Sequential Analysis* **31**(2), 238–262.

[3] De, S. and M. Baron (2012b). Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *J. Statist. Plann. Inference* **142**, 2059–2070.

[4] De, S. and M. Baron (2015). Sequential tests controlling generalized familywise error rates. *Statistical Methodology* **23**, 88–102.

# At-risk-of-poverty rate variance estimations using Gaussian kernel and smoothing splines in R package vardpoor

**Juris Breidaks**

*Central Statistical Bureau of Latvia, Latvia, juris.breidaks@csb.gov.lv*

**Keywords**: Smoothing splines, Gaussian kernel, variance estimation, vardpoor.

The Central Statistical Bureau of Latvia (CSB) in 2012 developed R [1] package vardpoor [2] (the set of functions for statistical calculation in programme R). The sampling error estimation mechanism of the R package vardpoor consists of the following procedures:

- calculation of the domain-specific study variables, if the sampling errors have to be estimated for population domains;
- at-risk-of-poverty rate (ARPR) [3] linearization using Gaussian kernel and smoothing splines [4];
- calculation of regression residual if the weights are calibrated;
- variance estimation with the ultimate cluster method [5].

The talk is devoted to estimation of the income distribution function using smoothing splines and Gaussian kernel. The linearized variables of ARPR are obtained using both linearization methods. Sampling errors are estimated for EU-SILC survey using both linearization methods.

## References

[1] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2018)

[2] Breidaks, J., Liberts, M., Ivanova, S.: vardpoor: Estimation of indicators on social exclusion and poverty and its linearization, variance estimation, R package version 0.9.11, Central Statistical Bureau of Latvia, Riga, Latvia, https://csblatvia.github.io/vardpoor/ (2017)

[3] Osier, G.: Variance estimation for complex indicators of poverty and inequality using linearization techniques. Survey Research Methods, 3, 167-195, http://w4.ub.uni-konstanz.de/srm/article/view/369 (2009)

[4] Asmuss, S., Breidaks, J., Budkina, N.: On approximation of density function by shape preserving smoothing histospline. Proceedings of the 15th Conference on Applied Mathematics (APLIMAT 2016), 30-43 (2016)

[5] Hansen, M., Hurwitz, W. N., Madow, W. G.: Sample Survey Methods and Theory I, Wiley, New-York (1953)

# Statistical challenges when analysing emerging epidemic outbreaks

**Tom Britton**[1]

[1] *Stockholm University, Sweden, tom.britton@math.su.se*

New infectious disease outbreaks have great impact on communities over the world, as recently manifested by the Ebola outbreak. An important statistical task is then to predict the future scenario with and with out preventive measures. In the current talk we will investigate such analyses and see how it can be improved. The main catch is that in the exponentially growing phase early on in an outbreak, several biases can occur if not taken account for: events with short delays will be over-represented. We will give some examples from the Ebola outbreak and see how the biases can be removed or at least reduced. (Joint work with Gianpaolo Scalia Tomba).

# Discrete state space models in survival analysis: a particle filter application

**Benjamin Christoffersen**[1]

[1] *Copenhagen Business School, Denmark, bch.fi@cbs.dk*

**Keywords**: Discrete hazard models, particle filtering, survival analysis, computational complexity, EM algorithm

I cover discrete hazard models with $i = 1, \dots, n_t$ individuals at risk at time $t = 1, \dots, d$ and latent state variables $\vec{\alpha}_1, \dots, \vec{\alpha}_d$. The dynamics are

$$\begin{aligned} \vec{\alpha}_t &= \mathbf{F}(\vec{\theta})\vec{\alpha}_{t-1} + \mathbf{R}\vec{\epsilon}_t & \vec{\epsilon}_t &\sim N(\vec{0}, \mathbf{Q}(\vec{\theta})) \\ & & \vec{\alpha}_0 &\sim N(\vec{0}, \mathbf{Q}_0) \\ \eta_{it} &= \vec{\beta}^\top \vec{x}_{it} + \vec{\alpha}_t^\top \mathbf{C}\vec{z}_{it} + o_{it} & y_{it} &\sim g(\eta_{it}) \end{aligned}$$

where $y_{it}$ is the observed survival or survival length in the period $(t-1, t]$, $\vec{x}_{it}$ and $\vec{z}_{it}$ are observation specific covariates, $o_{it}$ are offsets, and $\vec{\theta}$ and $\vec{\beta}$ are parameters to be estimated. I illustrate an application of particle filtering for estimating the parameters, $(\vec{\beta}, \vec{\theta})$, in an EM algorithm. The pros and cons of the particle filters are highlighted with an emphasis on the computational complexity.

## References

[1] Briers, M., Doucet, A., Maskell, S. (2009). Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics* **62**, no. 1, 61–89.

[2] Fearnhead, P., Wyncoll, D., Tawn, J. (2010). A Sequential Smoothing Algorithm with Linear Computational Cost. *Biometrika* **97**, no. 2, 447–64.

# Rare event simulation for GARCH(p,q) processes and related matrix recursive sequences

**Jeffrey F. Collamore**

[1] *University of Copenhagen, Denmark, collamore@math.ku.dk*

**Keywords**: Monte Carlo methods, importance sampling, random matrices, large deviations, risk theory, financial time series.

The primary goal of this talk is to present an efficient numerical method for estimating the tail probabilities in a general GARCH$(p,q)$ process. To this end, we develop an algorithm based on importance sampling, which applies not only to GARCH$(p,q)$ processes but, more generally, to recursive sequences of the form

$$V_n = A_n V_{n-1} + B_n, \quad n = 1, 2, \ldots,$$

where $\{(A_n, B_n) : n = 1, 2, \ldots\}$ are i.i.d. copies of $(A, B)$, $A$ is a $d \times d$ random matrix and $B$ a random vector in $\mathbb{R}^d$, both with nonnegative entries, and the matrix $A$ is "contracting" in the sense of [3], thus guaranteeing a unique stationary distribution for $\{V_n\}$. Our work builds upon [1], where we developed an analogous algorithm in the one-dimensional setting by introducing a novel "dual" change of measure to obtain efficiency. Similarly, our current work utilizes the regenerative structure of $\{V_n\}$ together with a generalization of this dual change of measure, suitably adapted to random matrices. (Based on joint work with A.N. Vidyashakar, G. Diao, and X. Xie in [2].)

## References

[1] Collamore, J.F., Diao, G., Vidyashankar, A.N. (2014). Rare event simulation for processes generated via stochastic fixed point equations. *Annals of Applied Probability* **24**, 2143-2175.

[2] Collamore, J.F., Vidyashankar, A.N., Diao, G., Xie, X. (2018). Rare event simulation for GARCH$(p,q)$ processes and related recursions generated by random matrices. Working paper.

[3] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* **131**, 207-248.

# II-CC-FF – combination of information beyond meta-analysis

**Céline Cunen**[1] **and Nils Lid Hjort**[2]

[1] *University of Oslo, Norway, cmlcunen@math.uio.no*
[2] *University of Oslo, Norway, nils@math.uio.no*

Combining information across different sources is an important statistical challenge, arising in many different fields. Combination is especially difficult when the sources are very diverse, requiring approaches beyond standard meta-analysis methods. The II-CC-FF paradigm (Independent Inspection, Confidence Conversion, Focused Fusion) is a general three-step method for such problems. The first step, II, uses different techniques to translate the information from each source to confidence distributions. Then these confidence distributions are transformed into confidence log-likelihoods in the CC step, before being combined in the FF step. In this talk, the II-CC-FF scheme will be presented and illustrated by several non-standard applications. Demonstrating that II-CC-FF can be applied successfully both in traditional meta-analysis settings and in more challenging cases. Guidelines, limitations and some theoretical results will also be discussed.

# On the equivalence between conditional and random-effects likelihoods in exponential families

**Riccardo De Bin**[1]

[1]*Department of Mathematics, University of Oslo, Norway, debin@math.uio.no*

**Keywords**: Modified profile likelihood; Neyman–Scott problem; stratified models.

The treatment of nuisance parameters is a central problem in statistical inference. When the number of nuisance parameters grows with the sample size, for example in case of stratified models with stratum-dependent nuisance parameters, some difficulties arise and the usual likelihood approaches are known to provide inconsistent results [2]. Among the solutions developed in the literature to face this issue, the conditional likelihood and the mixture models are arguably the most popular inferential tools. It has been noted that these two approaches provide very similar results in many situations [4], encouraging the investigation of their relationship and of the common properties.

Starting from Rice's contributions [3, 4] and focusing on the exponential family framework, here we show that the use of a particular mixing distribution based on moment generating functions leads to the equivalence between the random-effects (or marginal) likelihood used in the mixture models approach and the conditional likelihood. We also show that an approximation of the mixing function leads to the equivalence with the modified profile likelihood [1], a higher order asymptotics tool which may be used as an approximation of either marginal or conditional likelihoods when the latter do not exist.

## References

[1] Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

[2] Neyman, J., Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

[3] Rice, K. (2004). Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. *Journal of the American Statistical Association* **99**, 510–523.

[4] Rice, K. (2008). Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies. *Journal of the American Statistical Association* **103**, 385–396.

# Linear discriminant analysis of spatial Gaussian data

**Lina Dreižienė[1] and Kestutis Dučinskas[2]**

[1] *Vilnius University and Klaipėda University, Lithuania, l.dreiziene@gmail.com*
[2] *Vilnius University and Klaipėda University, Lihuania, k.ducinskas@gmail.com*

**Keywords**: Gaussian Markov random field; Bayes discriminant function; Actual and expected error rate.

This study is addressed to linear discriminant analysis of spatially correlated Gaussian data. Analytic expressions for the actual risk and the approximation of the expected risk based on the plug-in Bayes discriminant function are obtained. The aforementioned formulas are derived for two different types of Gaussian data: so-called geostatistical models with directly specified Matern type or other parametric covariance function models and Gaussian Markov random fields which are observed over regular or irregular lattices and specified by conditionally autoregressive model.

The accuracy of the proposed classifier is demonstrated using simulated data. Since the simulation study shows sufficiently high performance, application of the proposed classifier to the mapping of presence and absence of invasive species in the Curonian lagoon is presented.

# An Extension of the Fixed-Effects Principal Component Model to a Common Principal Component Environment

**Toni Duras**

*Jönköping University, Sweden, toni.duras@ju.se*

**Keywords**: Common Principal Component Analysis, Exploratory multivariate data analysis, Fixed-effects, Principal Component Analysis.

This paper is focused on exploring multivariate data using Principal Component Analysis (PCA). PCA can be conducted using different frameworks, all leading to the same mathematical base, the spectral analysis or the singular value decomposition. However, the results depend on the model or assumptions made. Some frameworks are very reluctant to assume probabilistic models, others are solely interested in distributional statement. Here, a model that lies in between the two extreme cases is considered, namely the fixed-effects PCA model. The model assumes the expectation of each data unit to lie in a lower dimensional space, the variance to define the noise in the data, and is estimated by minimizing a least squares criterium based on some selected metric. Finally, an attempt is made to generalize the fixed-effects PCA model to multiple populations when a common principal components structure is assumed.

# A class of asymptotically efficient estimators based on sample spacings

**M. Ekström[1], S. M. Mirakhmedov[2], and S. Rao Jammalamadaka[3]**

[1] *Umeå University, Sweden, Magnus.Ekstrom@umu.se*
[2] *Institute of Mathematics, National University of Uzbekistan, Uzbekistan, shmirakhmedov@yahoo.com*
[3] *University of California, Santa Barbara, USA, sreenivas@ucsb.edu*

**Keywords**: sample spacings, Csiszár divergences, parametric estimation, asymptotic efficiency, robustness.

We consider general classes of estimators based on higher-order sample spacings, called Generalized Spacings Estimators (GSEs). Such classes of estimators are obtained by minimizing approximations of Csiszár divergences between distributions in the model and the true underlying distribution; maximum likelihood estimators (MLEs) may be derived in a similar way using the Kullback-Leibler divergence. Our results generalize several earlier studies on spacings-based estimation, by utilizing non-overlapping spacings that are of an order which increases with the sample size. The GSEs are shown to be consistent as well as asymptotically normal under a fairly general set of regularity conditions. When both the order of the spacings and the number of spacings grow with the sample size, an asymptotically efficient class of estimators, called the "Minimum Power Divergence Estimators," are shown to exist. Simulation studies give further support to the performance of these asymptotically efficient estimators in finite samples, and compare well relative to the MLEs as well as corresponding estimators based on "overlapping" higher order spacings. Unlike the MLEs, some of these estimators are also shown to be quite robust under heavy contamination.

# Slowly converging sample means

**Benjamin Eltzner**

*University of Gottingen, Germany, beltzne@uni-goettingen.de*

The central limit theorem, which is among the foundations of statistics, says that the sample mean converges to the population mean with an asymptotic rate of $n^{1/2}$. There are numerous estimators known to statisticians which satisfy limit theorems with lower asymptotic rates than $n^{1/2}$. In such a case, hypothesis testing using asymptotic quantiles may require much larger samples. We show that a lower asymptotic rate for the mean can occur on manifolds of arbitrary finite dimension. We also highlight the impact of this fact on hypothesis testing especially in the case of high dimension and low sample size.

# Importance sampling and pseudomarginal Markov chain Monte Carlo methods for exact inference

**Jordan Franks**[1], **Matti Vihola**[1] **and Jouni Helske**[1,2]

[1] *University of Jyväskylä, Finland, jordan.j.franks@jyu.fi, matti.vihola@iki.fi*
[2] *University of Linköping, Campus Norrköping, Sweden, jouni.helske@iki.fi*

**Keywords**: Importance sampling, latent variable model, Markov chain Monte Carlo, parallel computing, pseudomarginal approach

We consider importance sampling (IS) type estimators [4], based on an approximate marginal Markov chain and subsequent IS weighting, and compare them with standard estimators based on the pseudomarginal (PM) approach [1, 2]. In the context of latent variable models, IS may be based on a Markov chain targeting an approximate marginal posterior, with IS weights formed based on unbiased estimators. We compare IS with PM and and its delayed acceptance (DA) variant in various experiments, which show that IS can be competitive even without parallelisation. This part of the talk is based on [4].

We also discuss theoretical comparison results of the IS and PM/DA approaches, based on [3]. We show asymptotic variance guarantees for IS in terms of PM/DA if the ratio of the true and approximate (marginal) likelihood is bounded. Often the ratio of the marginal likelihoods is bounded but the likelihood estimators are unbounded. We show that the IS chain is usually more stable than PM/DA in this case. Our experimental and theoretical results are promising, and point to the possible utility of the two-phase IS approach in various settings.

## References

[1] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342, 2010.

[2] C. Andrieu and G. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009.

[3] J. Franks and M. Vihola. Importance sampling correction versus standard averages of reversible MCMCs in terms of the asymptotic variance. Preprint arXiv:1706.09873, 2017.

[4] M. Vihola, J. Helske, and J. Franks. Importance sampling type estimators based on approximate marginal Markov chain Monte Carlo. Preprint arXiv:1609.02541, 2016.

# How to build confidence

**Kathrine Frey Frøslie**

*University of Oslo, Norway, k.f.froslie@medisin.uio.no*

Explaining the construction and interpretation of a confidence interval is a challenging and never-ending teaching task. In my talk I will present several teaching approaches that can make it easier to teach and learn confidence interval with confidence. Several statistical topics are used to view the confidence interval in different perspectives. These include variation vs estimation error, estimation error vs sample size, the Central Limit Theorem vs Bootstrapping, and a frequentistic vs Bayesian approach. Teaching material include cards made for classroom interaction, a manuscript for an effective blackboard-session, R simulations and visualizations.

# Bayesian Change-point Modelling of the Effects of 3-points-for-a-win Rule in Football

**Gebrenegus Ghilagaber and Parfait Munezero**

*Stockholm University, Sweden, Gebre@stat.su.se, Parfait.Munezero@stat.su.se*

We examine the effects of the 3-point-for-a-win (3pfaw) rule in the football (soccer) world. Data on mean goals and proportions of decided matches from seven leagues around the world form the basis of our analyses. Bayesian change-point analysis shows that the rule had no effects on the mean goals in any of the leagues but, indeed, had increasing effects on the proportion of decided matches in some of the leagues studied. This, in turn, implies that while the rule has given teams the incentive to aim at winning games, such aim was not achieved by scoring more goals but by scoring enough in order to win and, at the same, defending enough in order not to lose.

# A Fault Prediction Method for Temporal Data

**Emanuele Gramuglia**[1]

[1] *University of Oslo, Norway, emanueg@math.uio.no*

**Keywords**: Fault prediction, stochastic process, text categorization, pattern identification, features extraction.

Fault detection techniques have been studied by the statistics community since the late 1970s, encouraged by the fast development of software technologies for monitoring systems. The increased availability of complex and detailed information leads to several kinds of data structures involving time, such as data streams, temporal networks and time series data. Between these, event logs are becoming extensively used for monitoring because of their high reliability in determining the health status of the system. Our work provides a new log based statistical methodology for fault prediction. The model consists of two phases: pattern identification and feature extraction. For the first phase we assume an unobservable process of breakpoints defining patterns within the log file. The key feature for this process is its direct dependency on the observable series of events through functions such as the rate of occurrence of the events. Once the breakpoints are inferred, a new approach derived from the word space methodology is carried out in order to extract features. Such features represent the inputs to be used in several prediction methods such as penalised regression and neural nets.

# Functional approach to modelling of daily tax revenues

**Jovita Gudan**[1] **and Alfredas Račkauskas**[2]

[1] *Vilnius university, Lithuania, jovita.gudan@mif.vu.lt*
[2] *Vilnius university, Lithuania, alfredas.rackauskas@mif.vu.lt*

**Keywords**: functional data analysis, smoothing, registration, functional linear models.

We discuss functional data analysis approach to modelling and analysing daily tax revenues. The main features of daily tax revenue we need to extract are the patterns within calendar months. As standard seasonal time series techniques cannot be used due to varying number of banking days per calendar month and presence of seasonality between and within months we interpret monthly tax revenues as curves obtained from daily data. Standard smoothing techniques and registration of data according to some patterns within calendar months are used for data preparation. Our study can be described as an exploratory data approach: analysis $\Rightarrow$ model $\Rightarrow$ conclusions. Treating monthly time series as random functions in a space spanned by finite dimensional functional bases, we intensively explore methods of functional data analysis.

# From trees per hectare to nanoparticles: How Gibbs point processes model interaction in various scenarios

**Henrike Häbel**[1]

[1]*Natural Resources Institute Finland (Luke), Finland, henrike.habel@luke.fi*

**Keywords**: Forestry, Gibbs point processes, material structure, pair-potential function.

Spatial statistics has the nice feature that its methods can be applied to data on any scale ranging from nano- to kilometers and even beyond. Consequently, there are many applications where spatial statistics can be used for descriptive and modeling purposes. Here, the focus will be on (finite) pairwise interaction Gibbs point processes with applications in forestry and material sciences. In this context, points can be locations of trees or particles, for example. The points form point patterns in two or higher dimensions which are mathematically interpreted as realizations of a spatial point process.

Gibbs processes, sometimes also called Markov point processes, are a flexible family of point process models that allow for inter-point interactions with local or Markovian dependence between points. Pairwise attraction or repulsion can be described by a pair-potential function. Originating in statistical mechanics, the pair-potential function provides a link between physical and statistical models. This link becomes useful when using a static model for inference on a dynamic physical process underlying the formation of the observed point pattern.

In my talk, I will start with a rather typical application of Gibbs processes to regular forest [1]. Then I will show that Gibbs processes can also be used for modeling attraction in aggregated forests. I will, however, also discuss challenges faced when modelling aggregated structures on the example of a silica nanoparticle gel [2]. Last, but not least, I will present how Gibbs processes can even be applied to inhomogeneous and anisotropic porous films used for controlled drug release [3].

## References

[1] Stoyan, D., Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15** (1), 61–78.

[2] Häbel, H., Särkkä, A., Rudemo, M., Hamngren Blomqvist, C., Olsson, E., Abrahamsson, C., and Nordin, M. (2016). From static micrographs to particle aggregation dynamics in three dimensions. *Journal of Microscopy*, **262** (1):102–11.

[3] Häbel, H., Rajala, T., Boissier, C., Marucci, M., Schladitz, K., Redenbach, C., Särkkä, A. (2017). A three-dimensional anisotropic point process characterization for pharmaceutical coatings. *Spatial Statistics* **22** (2), 306–320.

# Post-selection inference: risk estimation after data-driven model selection

**Niels Richard Hansen, Frederik Riis Mikkelsen, Alexander Sokol**

*University of Copenhagen, Denmark, Niels.R.Hansen@math.ku.dk*

**Keywords**: Degrees of freedom, flows, model selection, post-selection inference, risk estimation, SURE.

Valid post-selection inference procedures are statistical methods that correctly account for data-driven model selection. AIC is an estimate of risk of the MLE using the log-likelihood loss, but it is an example of an invalid risk estimate under model selection. AIC as well as other standard information criteria, Stein's unbiased risk estimate (SURE), generalized cross validation and similar statistics are biased and they systematically underestimate risk under model selection.

Biased risk estimation is due to discontinuities of the underlying parameter estimators induced by the model selection algorithms. We have derived various representations and bounds, [1, 3], of the bias. In a special case of reduced rank regression it can be shown to vanish completely [1], but it is often of a considerable size.

In the talk a number of data-driven selection procedures including lasso-OLS, [1], marginal screening, forward step-wise selection and greedy basis pursuit will be shown to fit into a novel framework for risk estimation accounting for the model selection. In this framework, the selection events are related via a flow parametrized by a tuning parameter. In addition to valid risk estimates, our framework also provides a novel leverage measure for individual observations accounting for model selection.

## References

[1] Hansen, N.R. (2018). On Stein's unbiased risk estimate for reduced rank estimators. *Statistics & Probability Letters* **135**, 76-82.

[2] Mikkelsen, F.R. and Hansen, N.R. (2018). Degrees of freedom for piecewise Lipschitz estimators. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, to appear.

[3] Sokol, A. and Hansen, N.R. (2014). Degrees of freedom for nonlinear least squares estimation. *arXiv:1402.2997*.

# Peacekeeping Operations and the intensity of violence in internal armed conflicts

**Gudmund Hermansen**[1,2] **and Håvard Mokleiv Nygård**[3]

[1] *BI Norwegian Business School, Norway, gudmund.hermansen@bi.no*
[2] *University of Oslo, Norway*
[3] *The Peace Research Institute Oslo (PRIO), Norway, havnyg@prio.org*

**Keywords**: Bayes, event-level data, hierarchical modelling, Peacekeeping Operations.

We examine the effect of UN Peacekeeping Operations (PKOs) on the intensity of violence seen in civil wars. Reducing the intensity of violence is a core objective of PKOs but we still lack a proper understanding of how PKOs affect the underlying latent conflict-process. Using event-level data for all civil wars from 1989 to the present, we develop a Bayesian hierarchical modelling framework that allows us to study how deploying a PKO affects the intensity, i.e. the escalatory (and de-escalatory) patterns of violence, in these conflicts.

# Predictions and confidence

**Nils Lid Hjort**

*University of Oslo, Norway, nils@math.uio.no*

I will show how the machinery of confidence distributions and confidence curves can be used to provide inference for predictions. This includes themes like 'what will the next observation be like', estimating and providing uncertainty for a missing observation, forming a prediction band for the future of a time series, predicting an outcome based on a given covariate vector in a regression model, etc. Solutions are conveniently given as confidence curves for the not yet observed quantities. Illustrations for real data will be given.

## References

[1] Schweder, T. and Hjort, N.L. (2016). *Confidence, Likelihood, Probability*, Cambridge University Press.

# Localization of halfspace depth

**Daniel Hlubinka[1] Lukáš Kotík[2] and Ondřej Vencálek[3]**

[1] *Univerzita Karlova, Praha, Czechia, hlubinka@karlin.mff.cuni.cz*
[2] *National Radiation Protection Institute, Czechia*
[3] *Univerzita Palackého, Olomouc, Czechia*

**Keywords**: Data depth, localization.

Statistical depth functions are well-known nonparametric tools for analyzing multivariate data. Halfspace depth is most frequently used, and while it has many desirable properties, it is dependent on global characteristics of the underlying distribution. In some circumstances, however, it may be desirable to take into account more local and intrinsic characteristics of the data. To this end, we introduce weighted halfspace depths in which the indicator function of closed halfspace is replaced by a more general weight function. Our approach, which calls in part on functions associated with conic sections, encompasses as special cases the notions of sample halfspace depth and kernel density estimation. We give several illustrations and prove the strong uniform consistency of weighted halfspace depth incorporating mild conditions on the weight function.

## References

[1] Hlubinka, D., Kotík, L. and Vencálek, O. (2010). Weighted halfspace depth. *Kybernetika* **46**, 125–148.

[2] Kotík, L. and Hlubinka, D. (2017). A weighted localization of halfspace depth and its properties. *Journal of Multivariate Analysis* **157**, 53–69

# Hard functional parameter constraints in Bayesian estimation of economic models

**Øyvind Hoveid**

*Norwegian Institute of Bioeconomy Research*

**Keywords**: Bayesian estimation, hard constraints, economics.

Parameter constraints in Bayesian models are usually implemented softly with prior distributions. In effect, the likelihood is penalized more heavily the further the constraints are from holding. With a few constraints, heavy penalties may work well, but with many constraints and positive penalties, the likelihood may be distorted. With non-linear constraints the global minimum of joint penalties is not necessarily found — with additional harm to the likelihood. To avoid such problems I here turn to hard functional constraints of the form $F(\theta) = 0$, for some multidimensional continuous function $F$ defined on the parameter space of the model.

A sufficient requirement for modeling such constraints is a partition of parameters, $\theta = (\theta_1, \theta_2, \theta_3)$ with $\theta_i \in \Theta_i$, and for each $\theta_1$ a unique continuous 1-1 mapping $G_{\theta_1} : \Theta_2 \leftrightarrow \Theta_3$, defined implicitly by $F(\theta_1, \theta_2, G_{\theta_1}(\theta_2)) = 0$. In effect, the joint prior likelihood can be stated in two ways, as $\pi(\theta_1, \theta_2)$ or $\pi(\theta_1, \theta_3)$, related by the transformation $G_{\theta_1}$.

A convenient way of separating $\Theta_2$ and $\Theta_3$ is to let one sub-space, say $\Theta_2$, comprise the mostly observable parameters, and let the other comprise only non-observable ones. Simulations to find a sample from the posterior distribution can now take place within $(\Theta_1, \Theta_2)$. $\theta_3 \in \Theta_3$ can by assumption always be found with the function $G_{\theta_1}$. $G_{\theta_1}$ need not be explicitly defined. The implicit definition is sufficient as long as it is globally valid.

Economic models are to great extent stated in terms of equations — precisely of the form $F(\theta) = 0$ — representing first order and Euler conditions of optimality, equilibrium conditions, market balances etc.. On the one hand these equations represent parameter constraints based on theory which — if correct — sharpen the inference from the models. On the other hand, if incorrect, they will bias the inference. While frequentist models tend to be so sparsely parametrized that constraints to large extent are rejected, the case is more promising for Bayesian models. Richer parametrization leaves more relevance for constraints and a better frame for testing their relevance.

A case of partial disequilibrium among producers and consumers of meat is used for illustrations. The case is labeled partial because only meat prices are modeled, not the rest of the economy. Disequilibrium arises when the marginal value of traded meat vary systematically between producers and consumers. The objective of the analysis is to find the extent and the trends of this eventual asymmetry in a non-linear context — generalizing the linear frequentist model of [1] Breshnahan.

## References

[1] Bresnahan, Timothy F. (1982). The oligopoly solution concept is identified. *Economics Letters* **10** (1-2) 87–92.

# Deep Bayesian regression models

**Aliaksandr Hubin[1], Geir Storvik[2] and Florian Frommlet[3]**

[1]*University of Oslo, Norway, aliaksah@math.uio.no*
[2]*University of Oslo, Norway, geirs@math.uio.no*
[3]*Medical University of Vienna, Austria, florian.frommlet@meduniwien.ac.at*

**Keywords**: Bayesian deep learning, deep feature engineering, combinatorial optimization, uncertainty in deep learning, Bayesian model averaging, semi-parametric statistics, automatic neural network configuration, genetic algorithm, Monte Carlo Markov chains.

Regression models are addressed for inference and prediction in a wide range of applications providing a powerful scientific tool for the researchers and analysts coming from different fields. In most of these fields more and more sources of data are becoming available introducing a variety of hypothetical explanatory variables for these models to be considered. Model averaging induced by different combinations of these variables becomes extremely important for both good inference and prediction. Not less important, however, seems to be the quality of the set of explanatory variables to select from. It is often the case that linear relations between the explanatory variables and the response are not sufficient for the high quality inference or predictions. Introducing non-linearities and complex functional interactions based on the original explanatory variables can often significantly improve both predictive and inferential performance of the models. The non-linearities can be handled by deep learning models. These models, however, are often very difficult to specify and tune. Additionally they can often experience over-fitting issues. Random effects are also not incorporated in the existing deep learning approaches. In this paper we introduce a class of deep Bayesian regression models with latent Gaussian variables generalizing the classes of GLM, GLMM, ANN, CART, logic regressions and fractional polynomials into a powerful and flexible Bayesian framework. We then suggest algorithmic approaches for fitting them. In the experimental section we test some computational properties of the algorithm and show how deep Bayesian regression models can be used for inference and predictions in various applications.

# Detection of changes in time series of counts

**Šárka Hudecová[1], Marie Hušková[2] and Simos G. Meintanis[3]**

[1] *Charles University, Prague, Czech Republic, hudecova@karlin.mff.cuni.cz*
[2] *Charles University, Prague, Czech Republic, huskova@karlin.mff.cuni.cz*
[3] *National and Kapodistrian University of Athens, Greece, simosmei@econ.uoa.gr*

**Keywords**: change-point, integer autoregression model, Poisson autoregression.

We deal with a problem of sequential detection of structural changes in time series of discrete-valued observations, in particular in integer autoregression (INAR) and Poisson autoregression (Poisson INARCH) models. A change in the distribution as well as a parameter change are considered. The detector statistic is based on the empirical probability generating function and its asymptotic properties are studied under the null hypothesis as well as under alternatives. For practice, we propose a bootstrap test.

## References

[1] Hudecová, Š., Hušková, M., Meintanis, S.G. (2017). Tests for Structural Changes in Time Series of Counts. *Scandinavian Journal of Statistics* **44**, 843–865.

# Prior sensitivity quantification: To scale, or not to scale, that is the question

**Sona Hunanyan**[1] **and Małgorzata Roos**[1] **and Håvard Rue**[2]

[1] *University of Zurich, Switzerland, sona.hunanyan@uzh.ch, malgorzata.roos@uzh.ch*
[2] *King Abdullah University of Science and Technology, Saudi Arabia, haavard.rue@kaust.edu.sa*

**Keywords**: Formal sensitivity diagnostics, JAGS, Stan, R-INLA.

Bayesian hierarchical models (BHM) are nowadays a well established statistical methodology conveniently estimated by the Bayesian general-purpose software systems such as JAGS, Stan and R-INLA. BHMs have a unique ability to incorporate external prior knowledge in complex models. As prior input can vary considerably, sensitivity quantification is a crucial part of a careful Bayesian reasoning. Unfortunately, to-date a standardized sensitivity diagnostic tool for BHMs embracing all Bayesian software systems is missing.

The epsilon-local sensitivity measure introduced by [3] used prior neighbourhoods to provide a computationally cheap prior sensitivity assessment in BHMs computed by R-INLA. Neighbourhoods were also utilized by [4] for posterior distributions and by [1] for models. Unfortunately, neighbourhoods determination can be tedious and case-specific. In contrast, [2] suggested a scaling approach, which has a potential to completely dispense with neighbourhood considerations.

We will provide a proof of concept for the scaling sensitivity measure by comparing its feasibility with the epsilon-local one. A random-effects meta-analysis for historical trials used for prior elicitation in a medical study will exemplify our findings. We will discuss if the scaling sensitivity measure is a sound candidate for a unified and computationally cheap sensitivity diagnostic in BHMs fitted by R-INLA, JAGS and Stan.

## References

[1] Glad, I., Hjort, N.L. (2016). Model uncertainty first, not afterwards. *Statistical Science* **31**, 490–494.

[2] Guo, J. (2016). A study of local prior sensitivity of PC and inverse-gamma priors for variance parameters (Technical report). In: Guo, J, *Bivariate Meta-analysis*. PhD thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, 99–138.

[3] Roos, M., Martins, T.G., Held, L., Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis* **10**, 321–349.

[4] Watson, J., Holmes, Ch. (2016). Approximate models and robust decisions. *Statistical Science* **31**, 465–489.

# Inference in mixed models in R - beyond the usual asymptotic likelihood ratio test

**Ulrich Halekoh**[1] **and Søren Højsgaard**[2]

[1] *University of Southern Denmark, Denmark, uhalekoh@health.sdu.dk*
[2] *Aalborg University, Denmark, sorenh@math.aau.dk*

**Keywords**: Kenward-Roger approximation, parametric bootstrap, Satterthwaite approximation.

Mixed models in R (www.r-project.org) are currently usually handled with the `lme4` package. Until recently, inference (hypothesis test) in linear mixed models with `lme4` was commonly based on the limiting $\chi^2$ distribution of the likelihood ratio statistic. The `pbkrtest` package provides two alternatives: 1) A Kenward-Roger approximation for calculating (or estimating) the numerator degrees of freedom for an "F-like" test statistic. 2) $p$-values based on simulating the reference distribution of the likelihood ratio statistic via parametric bootstrap. A recent addition to the package is a Satterthwaite approximation of the degrees of freedom. We will illustrate the package through various examples, and discuss some directions for further developments.

## References

[1] Halekoh, U., Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models The R Package pbkrtest. *Journal of Statistical Software* **59**, 1–32.

# Bayesian networks with likelihood evidence in R

**Søren Højsgaard**[1]

[1] *Aalborg University, Denmark, sorenh@math.aau.dk*

**Keywords**: message passing, probability propagation.

The `gRain` package for R (www.r-project.org) has been around for a decade and the package provides tools for building Bayesian networks as well as for probability propagation in such networks. In `gRain`, attention is restricted to networks with discrete nodes.

A recent addition to the package is the ability to incorporate likelihood evidence (also called soft evidence or virtual evidence) in the network. This facility is useful in itself, and the facility also allows for handling networks where not all nodes are discrete.

## References

[1] Højsgaard, S. (2012). Graphical Independence Networks with the gRain Package for R. *Journal of Statistical Software* **46**, 1–26.

# On nonparametric depth based classification of functional observations

**Sami Helander[1], Stanislav Nagy[2], Germain van Bever[3], Lauri Viitasaari[4] and Pauliina Ilmonen[1]**

[1] *Aalto University School of Science, Finland, sami.helander@aalto.fi, pauliina.ilmonen@aalto.fi*
[2] *Charles University, Czech Republic, nagy@karlin.mff.cuni.cz*
[3] *Université libre de Bruxelles (ULB), Belgium, gvbever@ulb.ac.be*
[4] *University of Helsinki, Finland, lauri.viitasaari@iki.fi*

**Keywords**: functional data analysis, statistical depth, classification.

Lack of storage capacity and computing power limit us very little these days. We have been able to go from multivariate data to very high dimensional data. One approach used when dealing with such high dimensional data is to assume that the observed units are random functions (from some generating process) instead of random vectors.

The concept of statistical depth was originally introduced as a way to provide a nonparametric center-outward ordering from a depth-based multivariate median. Several different depth functions for functional observations have been presented in the literature. Most of these approaches, however, are based on assessing the location of the function as a measure of typicality. As a result, they are missing some important features inherent to functional data such as variation in shape. Another problem in assessing typicality of functional observations and in classifying functional observations is that we often observe only part of the function. One may overcome this problem by extrapolating and interpolating i.e. by adding the missing parts. However, doing that, at least implicitly, requires model assumptions.

We discuss assessing typicality of functional observations. Moreover, we provide a new classification method that is based on $j$-th order $k$-th moment integrated depths. For $j = 1$ and $k = 1$ this is equal to applying the mean halfspace depth of a functional value with respect to the corresponding univariate marginal distribution. When $j$ is larger than 1, the method is not based on comparing location only but considers shape of the function as well. Moreover, the method can be applied to partially observed functions without extrapolation or interpolation. Theoretical properties of the new approach are explored and several real data examples are presented to demonstrate its excellent classification performance.

# Fourth Cumulant for Multivariate Aggregate Claim Models

**Farrukh Javed[1], Nicola Loperfido[2] and Stepan Mazur[1]**

[1] *School of Business, Örebro University, Örebro, Sweden.*
[2] *University of Urbino, Urbino, Italy*

The fourth cumulant for the aggregated multivariate claims is considered in this article. We present a formula for the general case when the aggregating variable is independent of the multivariate claims. Two important special cases are also considered. In the first one, multivariate skewed normal claims are considered and aggregated by a Poisson variable. The second case is dealing with multivariate asymmetric generalized Laplace and aggregation is made by a negative binomial variable. Due to the invariance property, the latter case can be derived directly, leading to the identity involving the cumulant of the claims and the aggregated claims. There is a well established relation between asymmetric Laplace motion and negative binomial process that corresponds to the invariance principle of the aggregating claims for the generalized asymmetric Laplace distribution. We explore this relationship and provide multivariate continuous time version of the results. It is discussed how these results that deal only with dependence in the claim sizes can be used to obtain a formula for the fourth cumulant for more complex aggregate models of multivariate claims in which the dependence is also in the aggregating variables.

# Identifying and predicting fluid build-up in the lungs

**Andreas Kryger Jensen**

*University of Copenhagen, Denmark,*
*aeje@sund.ku.dk*

I consider the clinical problem of identifying and predicting fluid build-up in the lungs as a consequence of reduced heart function. The physical composition of a lung can though computed tomography (CT) scanning be summarized by a density function. The shape of such a lung density is presumed to change according to amount of fluid present in the lungs.

Densities are a type of functional data and therefore being of infinite dimensionality. Furthermore, density functions belong to a non-linear manifold due to their positivity and integral constraint. To accommodate this non-linearity we suggest a transformation approach based on the Riemannian geometry of the Hilbert sphere. The statistical modeling can then be performed in a suitably chosen tangent space of the sphere.

I will present both a non-parametric and a semi-parametric regression model where the functional objects are random density functions. In the non-parametric model the densities are functional covariates and the stage of heart function is the outcome. For the semi-parametric model the density functions are the outcome. In this model we approximate the densities in a finite dimensional sub-manifold generated by a parametric reference family and the residuals are smooth diffeomorphisms.

I illustrate the applications of the models on a data set obtained from CT scannings of pigs with induced lung fluid build-up. The analyses show that the regression models are capable of both predicting the degree of fluid build-up as well as describing the changes in the shape of the densities enforced by the amount of fluid.

# Lambert $W$ random variables as a tool for skewing distributions

**Meelis Kääärik**[1], **Tuuli Puhkim**[1] **and Liivika Tee**[1]

[1] *University of Tartu, Estonia, meelis.kaarik@ut.ee*

**Keywords**: skewed distributions, Lambert $W$ random variables, loss distributions, non-life insurance

There are different tools to assess skewed data, several skewed families are proposed, think, e.g., of skew-normal and related distributions. The Lambert $W$ random variables approach is an alternative where certain transform is proposed for a known random variable rather than creating of a new one[1]. Thus, we can construct a Lambert $W$ skewed version from any distribution. Since the Lambert $W$ function is double-valued, we distinguish the corresponding branches. Both principal and non-principal branches are analyzed theoretically. The values of the skewness parameters leading to the extreme values of the Lambert $W$ function are also investigated. In the practical part the suitability of corresponding location-scale distributions as well as Lambert $W$ transformed exponential distribution are evaluated on real insurance data. The results are compared with common loss distributions.

## References

[1] G. M. Goerg (2011). *Lambert W Random Variables - A New Family of Generalized Skewed Distributions with Applications to Risk Estimation*. Annals of Applied Statistics, **3** (5), 2197-2230.

# Parameter estimators of sparse random intersection graphs with thinned communities

**Joona Karjalainen[1], Johan S.H. van Leeuwaarden[2] and Lasse Leskelä[3]**

[1] *Aalto University, Finland, joona.karjalainen@aalto.fi*
[2] *Eindhoven University of Technology, The Netherlands, j.s.h.v.leeuwaarden@tue.nl*
[3] *Aalto University, Finland, lasse.leskela@aalto.fi*

**Keywords**: statistical network analysis, random graphs, parameter estimation, overlapping communities, partial data

We study a statistical network model generated by a large number of randomly sized overlapping communities, where any pair of nodes sharing a community is linked with probability $q$ via the community.

In the special case with $q = 1$ the model reduces to a random intersection graph which is known to generate high levels of transitivity also in the sparse context. The parameter $q$ adds a degree of freedom and leads to a parsimonious and analytically tractable network model with tunable density, transitivity, and degree fluctuations.

We prove that the parameters of this model can be consistently estimated in the large and sparse limiting regime using moment estimators based on partially observed densities of links, 2-stars, and triangles.

## References

[1] Karjalainen, J., van Leeuwaarden, J.S.H., Leskelä, L. (2018). Parameter estimators of sparse random intersection graphs with thinned communities. *arXiv:1802.01171*. `https://arxiv.org/abs/1802.01171`

# Towards automated causal inference

**Juha Karvanen**

*University of Jyväskylä*

**Keywords**: algorithm, causality, graphical model, identifiability, pruning, simplification

Causal inference refers to drawing conclusions on the effects of causes on the basis of experimental and observational data and expert knowledge. Many parts of this process can be automated. In our vision, the conclusions are derived by user-friendly algorithms after the researcher has specified the causal structure, the study design and the missing data mechanism.

Many building blocks for this vision exist already. One of the great achievements of the field is the derivation of algorithms for the identification of causal effects from observational data [1, 2]. These algorithms and their extensions to selection bias and transportability problems [3] have been implemented in the R package `causaleffect` [4]. To improve the practical usability of these algorithms, we have studied automated simplification of the resulting probabilistic expressions [5, 6].

In an ongoing work, we aim to build tools for problems that are unattainable by the current algorithms. As an example, we consider identification of causal effects by combined use of experimental and observational data.

## References

[1] Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573. AAAI Press/The MIT Press.

[2] Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1219–1226. AAAI Press.

[3] Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595.

[4] Tikka, S. and Karvanen, J. (2017). Identifying causal effects with the R package causaleffect. *Journal of Statistical Software*, 76.

[5] Tikka, S. and Karvanen, J. (2017). Simplifying probabilistic expression in causal inference. *Journal of Machine Learning Research*, 18(36):1–30.

[6] Tikka, S. and Karvanen, J. (2017). Enhancing identification of causal effects by pruning. *Submitted*.

# Network models and sparse graphon estimation

**Olga Klopp**[1], **Alexandre Tsybakov**[2] **and Nicolas Verzelen**[3]

[1] *ESSEC Business School and CREST, France, kloppolga@math.cnrs.fr*
[2] *ENSAE, France, alexandre.tsybakov@ensae.fr*
[3] *INRA, France, nicolas.verzelen@inra.fr*

**Keywords**: inhomogeneous random graph, networks, oracle inequality, sparse graphon, sparsity, stochastic block model

Inhomogeneous random graph models encompass many network models such as stochastic block models and latent position models. We consider the problem of statistical estimation of the matrix of connection probabilities based on the observations of the adjacency matrix of the network. Taking the stochastic block model as an approximation, we construct estimators of network connection probabilities – the ordinary block constant least squares estimator, and its restricted version. We show that they satisfy oracle inequalities with respect to the block constant oracle. As a consequence, we derive optimal rates of estimation of the probability matrix. Our results cover the important setting of sparse networks. Another consequence consists in establishing upper bounds on the minimax risks for graphon estimation in the $L_2$ norm when the probability matrix is sampled according to a graphon model. These bounds include an additional term accounting for the "agnostic" error induced by the variability of the latent unobserved variables of the graphon model. In this setting, the optimal rates are influenced not only by the bias and variance components as in usual nonparametric problems but also include the third component, which is the agnostic error. The results shed light on the differences between estimation under the empirical loss (the probability matrix estimation) and under the integrated loss (the graphon estimation).

# Focused information criterion for copula

**Vinnie Ko**[1] **and Nils Lid Hjort**[2]

[1] *University of Oslo, Norway, vinniebk@math.uio.no*
[2] *University of Oslo, Norway, nils@math.uio.no*

**Keywords**: copula, model selection, focused information criterion, pseudo maximum likelihood, two-stage maximum likelihood, inference functions for margins.

Focused information criterion (FIC) [1] is a model selection criterion that can compare non-nested parametric models with a non-parametric alternative. FIC evaluates candidate models based on the estimated precision of the focus parameter, where focus parameter is a population quantity of interest in statistical analysis (e.g. upper-tail probability and copula parameter).

By extending the FIC theories of [1], we develop the focused information criterion (FIC) for copula models. The FIC for copula models covers the three most popular estimation schemes: maximum likelihood, two-stage maximum likelihood (also known as inference functions for margins) and pseudo maximum likelihood.

By using the FIC for copula, one can compare the performance of models that are estimated with different estimation schemes. This is one of the biggest advantage of the FIC and is not possible with traditional model selection criteria like AIC and BIC, since they evaluate the model in terms of Kullback–Leibler divergence from the data generating model.

We demonstrate the behavior of FIC by using a simulation study and via an application to real data.

## References

[1] Jullum, M. & Hjort, N. L. (2017). Parametric or nonparametric: the fic approach. *Statistica Sinica* **27**, 951–981.

# A journey from multivariate normality to skewed data models

**Tõnu Kollo**

*University of Tartu, Estonia, tonu.kollo@ut.ee*

**Keywords**: copula, density approximation, elliptical distribution, multivariate normal distribution, skew elliptical distribution.

In "Introduction to the Multivariate Statistical Analysis", 1958, T. W. Anderson presented statistical models for the normal population in a compact matrix form. This became a dominating model in data analysis for years.

*Elliptical distributions.* As a generalization of the normal distribution Kelker (1970) introduced elliptical distributions. The first monographs on the topic appeared twenty years later. As in the normal case the distributions were elliptically contoured and axial symmetric.

*Density approximations.* Cornish & Fisher (1937) presented a complicated density function of $Y$ through a simple density of $X$ as a series expansion where moments of both distributions and derivatives of the density of $X$ were involved. Finney (1963) carried over the idea to the multivariate case and gave a general expression between two densities in $R^p$. The approach made it possible to tranform a symmetric (normal) density to skewed density approximation. I. Traat presented the multivariate density approximation in matrix representation in 1986. A general relation between two different dimensional densities was derived in 1998 by T. Kollo and D. von Rosen.

*Skew elliptical distributions.* In 1996 multivariate skew normal distribution was introduced by Adelchi Azzalini. The construction of transforming symmetric normal distribution to a skewed distribution was soon carried over to other continuous elliptical distributions. The distributions became quickly popular in modelling skewed data.

*Copulas.* Multivariate density approximations and skew elliptical distributions have marginals from the same distribution families (normal, $t$-distributed, for instance). Copula is a multivariate distribution function with uniform marginals and given dependence structure. Copula models enable to construct skewed data models with different type of marginals and given correlation structure. Nelsen (1998) gives the first monographic overview of the topic.

## References

[1] Cornish, E.A. & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *International Statistical Review* **5**, 307-322.

[2] Finney, D.J. (1963). Some properties of a distribution specified by its cumulants. *Technometrics* **5**, 63–69. C

[3] Nelsen, R. (1998). *An Introduction to Copulas.* Springer, New York.

# Missing data and the maximum contrastive pessimistic likelihood

**Jesse H. Krijthe**[1], **Marco Loog**[2,3], **Bo Markussen**[3]

[1] *Radboud University Nijmegen, The Netherlands*
[2] *Delft University of Technology, The Netherlands*
[3] *University of Copenhagen, Denmark*

**Keywords**: log-likelihood estimation, missing data, contrast, pessimism.

We consider log-likelihood estimation in the missing data setting for parametric models $p(\cdot|\theta)$, studying the particular estimator that solves the maximin problem

$$\operatorname*{argmax}_{\theta\in\Theta} \sum_{i=1}^{N} \min_{P\in\mathcal{P}_i} \int \log p((y_i^{\text{obs}}, y_i^{\text{mis}})|\theta) - \log p((y_i^{\text{obs}}, y_i^{\text{mis}})|\hat{\theta}_{\text{CC}})\, dP(y_i^{\text{mis}}).$$

Here $\Theta$ is the parameter space, $\hat{\theta}_{\text{CC}} \in \Theta$ is the complete-case ML estimate, $N$ is the number of data points, $y_i^{\text{obs}}$ the observed part of every data point, and $y_i^{\text{mis}}$ is the missing part. To keep the notation light, it is not made explicit that the specific variables observed and missing are not the same for every index $i$. Finally, $\mathcal{P}_i$ is the space of all probability measures $P$ on the domain of missing variables $y_i^{\text{mis}}$.

Our estimator generalizes the so-called maximum contrastive pessimistic likelihood (MCPL) estimator from semi-supervised learning [1]. Semi-supervised learning is concerned with ways to improve standard supervised classification using additional unlabeled data. When discrete labels are missing, the above integrals reduce to sums over $y_i^{\text{mis}}$ with $dP(y_i^{\text{mis}})$ the probability of associating label $y_i^{\text{mis}}$ to $y_i^{\text{obs}}$.

By subtraction, our formulation *contrasts* the log-likelihood of $\hat{\theta}_{\text{CC}}$ explicitly with that of the generalized estimate over all data, including everything missing. Additionally, it considers the *pessimistic* worst-case behavior for the probability measures on the missing data. This ensures the MCPL estimate to not perform worse than $\hat{\theta}_{\text{CC}}$ in terms of the log-likelihood on the full data, regardless of the true $y_i^{\text{mis}}$. Note that EM, the *de facto* standard, does not come with such guarantee.

For some classical discriminant analyses, MCPL actually provides demonstrably improved log-likelihoods under semi-supervision [1]. So when does generalized MCPL improve over the complete-case? One complication answering this arises when the missing variables space is not finite. We will revisit semi-supervised LDA [1], discuss logistic regression in that setting [2], consider estimating the parameters of a normal distribution, and study some further, less standard models.

## References

[1] Loog, M. (2016). Contrastive pessimistic likelihood estimation for semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38(3)**, 462–475.

[2] Krijthe, J.H., Loog, M. (2016). The pessimistic limits of margin-based losses in semi-supervised learning. *arXiv* **1612.08875**.

# Rate optimal estimation of quadratic functionals in inverse problems with partially known operator

**Martin Kroll**[1]

[1] *École Nationale de la Statistique et de l'Administration Économique, France,*
*martin.kroll@ensae.fr*

**Keywords**: Gaussian sequence model, quadratic functional, minimax theory, nonparametric testing.

We consider the Gaussian sequence model given by

$$\begin{aligned} X_j &= \lambda_j \theta_j + \varepsilon \xi_j, \quad j \in \mathbb{N}, \qquad \text{and} \\ Y_j &= \lambda_j + \sigma \eta_j, \quad j \in \mathbb{N}, \end{aligned}$$

where the parameter $\theta = (\theta_j)_{j\in\mathbb{N}}$ belongs to some smoothness class $\Theta \subseteq \ell^2(\mathbb{N})$, $(\xi_j)_{j\in\mathbb{N}}$ and $(\eta_j)_{j\in\mathbb{N}}$ are independent random vectors with i.i.d. $\sim \mathcal{N}(0,1)$ distributed components, and $\varepsilon, \sigma \in (0,1)$ are the known noise levels. We consider the estimation of quadratic functionals of the form

$$\mathfrak{q}(\theta) = \sum_{j=1}^{\infty} \omega_j^2 (\theta_j - \theta_j^\circ)^2$$

where $(\omega_j)_{j\in\mathbb{N}}$ is a weight sequence and $(\theta_j^\circ)_{j\in\mathbb{N}}$ a fixed element in $\Theta$. Note that the estimation of quadratic functionals in inverse problems has already been investigated in [1] where upper risk bounds were established. In contrast to [1], we dispense with the classical assumption that the sequence $\lambda = (\lambda_j)_{j\in\mathbb{N}}$ is known but assume that it is a nuisance parameter that can only be accessed by means of the additional observations $(Y_j)_{j\in\mathbb{N}}$. Under smoothness assumptions on the sequences $\theta$ and $\lambda$ we derive the optimal rates of convergence for the estimation of $\mathfrak{q}(\theta)$. In particular, we show that a truncated series estimator of the form

$$\widehat{\mathfrak{q}}_k = \sum_{j=1}^{k} \omega_j^2 \frac{X_j^2 - \varepsilon^2}{Y_j^2 - \sigma^2} \mathbf{1}_{\{Y_j^2 \geq 2\sigma^2\}} - 2 \sum_{j=1}^{k} \omega_j^2 \theta_j^\circ \frac{X_j}{Y_j} \mathbf{1}_{\{Y_j^2 \geq 2\sigma^2\}} + \sum_{j=1}^{\infty} \omega_j^2 (\theta_j^\circ)^2$$

with a suitably chosen value of $k \in \mathbb{N}$ can attain the optimal rate of convergence. Finally, we discuss implications of the derived results for the theory of signal detection and goodness-of-fit testing in inverse problems. Some potential directions for future research will be indicated to conclude the talk.

## References

[1] Butucea, C., Meziani, K. (2011). Quadratic functional estimation in inverse problems. *Statistical Methodology* **8**, 31–41.

# On Empirical Cumulant Function Based Estimation in Stable Laws

**Annika Krutto**[1], **Tõnu Kollo**[1]

[1] *University of Tartu, Estonia*

**Keywords**: asymptotic properties, empirical characteristic function, covariance matrix, point estimation, Monte-Carlo simulations.

The four-parameter stable distributions are adaptable statistical distributions that can capture the fuzzy dynamics and large fluctuations often seen in various processes in science and engineering. However, a challenging problem in applying stable distributions to practical problems is estimating their parameters because many stable distributions have densities with no explicit form and infinite moments. Popular parameter estimation methods for general stable distributions are algorithmic procedures rather than closed form estimators. To address this problem, [1] studied a class of closed-form estimators, called cumulant estimators, which provides statistically consistent estimators with no restriction on the parameter space. Cumulant estimators are based on the logarithm of empirical characteristic function at only two distinct positive real arguments, and have great computational simplicity. Although [1] performed an empirical search on the selection of arguments, no formal solution was provided. This paper extends [1] in two directions: (i) it proves that cumulant estimators have asymptotic normality and (ii) it provides a general rule for selecting the two arguments. Our extensive simulations show that under the provided selection rule, the closed-form cumulant estimators generally out-perform the more well-known algorithmic methods.

## References

[1] Krutto, A. (2016). Parameter Estimation in Stable Law. *Risks* **4**, 4 49.

# Bayesian inference for tree regeneration models using MCMC and INLA

**Mikko Kuronen**[1], **Mari Myllymäki**[2] **and Matti Vihola**[3]

[1] *Natural Resources Institute Finland (LUKE), Finland, mikko.kuronen@luke.fi*
[2] *Natural Resources Institute Finland (LUKE), Finland, mari.myllymaki@luke.fi*
[3] *University of Jyväskylä, Finland, matti.vihola@iki.fi*

**Keywords**: spatial random effects, approximate Bayesian inference, competition kernel.

We model tree regeration including birth and growth using a non-linear mixed model with spatial correlated random effects. Fixed predictors in the model are related to tree size and competition. A key component of interest is the effect of nearby larger trees. The spatial random effects are modeled using a Gaussian process with Matern covariance function. We perform a Bayesian inference for the models. The random effects part of the model is approximated using INLA, while inference for the hyper parameters of the random effects as well as the fixed effects part is done using MCMC.

# Stochastic optimization program with random decision points

**Petr Lachout**[1]

[1] *Charles University, Prague, Czech Republic, Petr.Lachout@mff.cuni.cz*

**Keywords**: Stochastic optimization program, random decision points, nodes of a graph.

We consider a random process acting on nodes of a given graph. Nodes are activated in random times and considered process is controlled in nodes which are active. Inactive nodes are unreachable for the process. The task is to optimize gain from the controlled random process dynamically developing in time. We intend to introduce a description of the random optimization scheme and present some basic relations, results.

## References

[1] Bonnans J.F., Shapiro A. (2000). *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York.

[2] Rockafellar R.T., Wets R. J.-B. (1998). *Variational Analysis*. Springer-Verlag, Berlin.

[3] Shapiro A., Dentcheva D., Ruszczynski A. (2009). *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM, Philadelphia.

# Confidence intervals for ranks

**Sture Holm [1], Rolf Larsson [2]**

[1] *Chalmers University of Technology, holm@chalmers.se*
[2] *Uppsala University, rolf.larsson@math.uu.se*

A common problem with ranking lists (e.g. regarding success/failiure proportions of treatments at hospitals) is that smaller units tend to end up either in the top or in the bottom just by pure chance. To alleviate this problem, we propose a method that, for a given unit, gives a confidence interval for the position of this unit within the distribution of the other units. The confidence interval is based on asymptotic normality. The method is illustrated by an empirical example, ranking hospitals in Sweden w.r.t. mortality during treatment after heart attack.

# Cauchy difference priors for edge-preserving Bayesian estimation in inverse problems

**Sari Lasanen[1], Lassi Roininen [1], Markku Markkanen[2] and Janne M.J. Huttunen [3]**

[1] *University of Oulu, Finland, sari.lasanen@oulu.fi, lassi.roininen@oulu.fi*
[2] *Eigenor Corporation, Finland, markku.markkane@eigenor.com*
[3] *Nokia Corporation, janne.m.huttunen@nokia.com*

**Keywords**: Bayesian statistical inverse problems, Cauchy distribution, Lévy $\alpha$-stable motion

We consider estimation problems in which the unknown includes sharp edges, for example interfaces between different materials. Such problems are typical in image reconstruction, tomography, and other inverse problems algorithms. A common solution for edge-preserving inversion is to use total variation (TV) priors. However, as shown by Lassas and Siltanen [1], TV-prior is not discretization-invariant: the edge-preserving property is lost when the computational mesh is made denser and denser. We propose another class of priors for edge-preserving Bayesian inversion, the Cauchy difference priors. We construct Cauchy priors starting from continuous one-dimensional Cauchy motion, and show that its discretized version, Cauchy random walk, can be used as a non-Gaussian prior for edge-preserving Bayesian inversion. We generalize the methodology to two-dimensional Cauchy fields, and briefly consider a generalization of the Cauchy priors to Lévy $\alpha$-stable random field priors. We develop a suitable posterior distribution sampling algorithm for conditional mean estimates with single-component Metropolis-Hastings.

## References

[1] M. Lassas and S. Siltanen, Can one use total variation prior for edge preserving Bayesian inversion?, Inverse Problems 20 (2004), 1537–1564.

# Maximum likelihood estimation in Gaussian distributions under total positivity

**Steffen Lauritzen**

*University of Copenhagen, lauritzen@math.ku.dk*

**Keywords**: Attractive Gaussian Markov random field, Gaussian graphical model; inverse M-matrix; ultrametric.

The problem of maximum likelihood estimation for Gaussian distributions that are multivariate totally positive of order two (MTP2) is investigated. The maximum likelihood estimator (MLE) for such distributions exists based on just two observations, irrespective of the underlying dimension. It is further demonstrated that the MTP2 constraint serves as an implicit regularizer and leads to sparsity in the estimated inverse covariance matrix, determining what we name the ML graph. It is shown that the maximum weight spanning forest (MWSF) of the empirical correlation matrix is a spanning forest of the ML graph. In addition, an upper bound for the ML graph is found by adding edges to the MSWF corresponding to correlations in excess of those explained by the forest. Globally convergent coordinate descent algorithms for calculating the MLE under the MTP2 constraint are provided, structurally similar to iterative proportional scaling. The lecture is based on joint work with Caroline Uhler and Piotr Zwiernik.

## References

[1] Lauritzen, S., Uhler, C., and Zwiernik, P. (2017). Maximum likelihood estimation in Gaussian models under total positivity. To appear in *The Annals of Statistics*. arXiv:1702.04031.

# Valuation of stochastic cash flows subject to capital requirements

**Filip Lindskog**[1]

[1] *Stockholm University, Sweden, lindskog@math.su.se*

**Keywords**: multi-period valuation, dynamic risk measure, liability cash flow.

Banks and insurance companies have contracted obligations to make future payments. For insurance companies these payments are to a large extent economic compensation for accidents suffered by current and past customers. In order to safeguard that these payments can be made, capital requirements are imposed. On the one hand, a capital requirement at a given time depends on the value of the liability cash flow at a later time. On the other hand, a value at a given time of the liability cash flow depends on future unknown capital requirements. I will present the basic ingredients of mathematical frameworks for valuing stochastic cash flows subject to capital requirements, including dynamic risk measurement and properties and computational aspects of the backward recursions that appear and define the stochastic process of cash flow values.

Capital requirements are imposed at a predefined set of times leading to discrete-time value processes defined on this time grid. If the distance between these times is small, then we may study continuous-time value processes that appear when considering a sequence of partitions of the entire time period whose meshes tend to 0. Existence, properties and interpretations of continuous-time limits of discrete-time value processes will be discussed.

The talk is primarily based on [1], [2], [3].

## References

[1] Engsner, H., Lindholm, M., Lindskog, F. (2017). Insurance valuation: A multi-period cost-of-capital approach. *Insurance: Mathematics and Economics* **72**, 250–264.

[2] Engsner, H., Lindensjö, K., Lindskog, F. (2017). The value of a liability cash flow in discrete time subject to capital requirements. *Submitted*

[3] Engsner, H., Lindskog, F. (2018). Continuous-time limits of multi-period cost-of-capital valuations. *Submitted*

# Spatial Reconstruction of Forest Fires

**Johan Lindström[1], Julia Hjalmarsson[2], Richard Bradshaw[3], Bengt Gunnar Jonsson[2] and Veiko Lehsten[1,4]**

[1]*Lund University, Sweden, johanl@maths.lth.se;* [2]*Mid Sweden University, Sweden;* [3]*University of Liverpool, United Kingdom;* [4]*Swiss Federal Insitute for Forest, Switzerland*

**Keywords**: Spatial statistics, Generalized linear mixed, logistic regression, Gaussian Markov random fields, Forest fires.

Forest fires constitute a major source of ecosystem disturbances. Their impact on ecosystem properties such as biodiversity has been acknowledged and included in certification standards for sustainable forestry; and restoration fires are increasingly used in active management of protected areas. While the extent of modern forest fires can be estimated from satellite data the time period covered by these data is very short compared to the time scale at which forest fires typically re-occur; 50–150 years for Northern boreal forests. Scandinavian and Siberian forests mainly experience ground fires that have a low probability of killing the trees, instead leaving fire scars which can been used to reconstruct the extent of past fires

Here we introduce a generalized linear model based on logistic regression with an underlying Gaussian Markov random field (GMRF) [2] that captures the spatial dependence between burnt trees. Formulating the spatial dependence as a GMRF allows the "distance" between trees to be defined through a triangulation of the study area; thus lakes and coastlines can be included to naturally restrict the fire reconstructions to go around larger lakes instead of jumping across. The resulting model was fitted using the R-INLA package [1]. Framing the model in a form that allows for the use of standard statistical software was done to increases the model's disseminate among non-statisticians.

The suggested method was evaluated on historic data from the Jämtgaveln Nature Reserve in Central Sweden and the results were compared to reconstructions from conventional

methods such as Thiessen polygons, grid cells containing burnt trees, and manual delineation. A large number of landscape features have traditionally been used to inform the manual delineation. These features can be included as covariates in the statistical model and the analysis indicated that few of these features are significant.

## References

[1] Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *J. Stat. Softw.*, 63(19):1–25.

[2] Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. B*, 73(4):423–498.

# Estimation for Dynamical Spatio-temporal Array Models

**Adam Lund**[1]

[1] *University of Copenhagen, Denmark, adam.lund@math.ku.dk*

**Keywords**: VSD imaging data, stochastic functional differential equation, non-differentiable regularization, GLAM

We present a computationally attractive framework for dynamical modeling of spatio-temporal data sampled in a regular grid (array data). In this framework a stochastic functional differential equation (SFDE) is proposed as a continuous space-time model for the dynamics of the random field underlying the data. The drift operator in this SFDE is decomposed into a sum containing a longe range memory component (a linear casual filter) and a (deterministic) transient component.

The framework establishes a generative model (the SFDE) for spatio-temporal data as a computationally feasible statistical model for realistic large scale settings. This is achieved by exploiting the array structure to represent the drift components using tensor product basis functions.

We apply this framework to a neuronal brain imaging data set to extract structure from data – smooth functions – that quantify components related to the propagation of brain activity occurring as a response to a visual stimulus. In particular by fitting the impulse response function in the linear casual filter to data we obtain an estimate of the synaptic weight function, which in turn quantify functional brain connectivity.

## References

[1] Adam Lund, Niels Richard Hansen (2018). Sparse Network Estimation for Dynamical Spatio-temporal Array Models. *ArXiv*.

# Index-based methodology in population statistics

**Ethel Maasing [1], Ene-Margit Tiit [2], Helle Visk [3] and Mare Vähi[4]**

[1] *Statistics Estonia, ethel.maasing@stat.ee*
[2] *Statistics Estonia, University of Tartu, ene.tiit@ut.ee*
[3] *Statistics Estonia, helle.visk@stat.ee*
[4] *University of Tartu, Estonia, mare.vahi@ut.ee*

**Keywords**: Population, residency index, partnership index .

Using registers instead of surveys and censuses is the new approach in official statistics[1]. But using registers in statistics is possible if registers are correct. Errors and biases are most typical in registers that are connected with some bonuses, for instance – distribution of children by schools and kindergartens by their addresses in Population Register causes the situation that children (and at least one parent) their address in the area of wished school or kindergarten.

In Statistics Estonia the methodology has been created to check the validity of Population Register data and improving their quality using other administrative registers and data sources.

The first task solved was annual estimation of population size or the number of residents. To check annually if the person is a resident or has left the country, the residency index has been created that can be interpreted as the person's probability of being resident [2]. The person has been counted to the set of residents if and only if the value of the index, calculated as linear combination of administrative signs of life (signs of activity in some register) is higher than empirically fixed threshold. This index is in use in Statistics Estonia for estimating the population size, yearly immigration and emigration rates. It also has some perspective for estimating trans-national population. The next step is to calculate the partnership index for statistical deciding if a pair of persons forms a family nucleus [3]. In this area only the first solution has been made and the results will be checked using a special survey. The following step will be finding the right address for people who have registered an incorrect address in the Population Register. Potential new sources for location data include big data, e.g. electricity consumption.

## References

[1] Tiit,E.-M., Vähi, M. (2017). Indexes in demographic statistics: a methodology using nonstandard information for solving critical problems. *Papers on Anthropology* **XXVI/1**, 72-87

[2] Maasing, E., Tiit, E.-M., Vähi, M. (2017). Residency index – a tool for measuring the population size. *ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA* **Volume 21, Number 1**, 129-139.

[3] Tiit, E.-M., Visk,H., Levenko, V. (2018). Partnership index. **Quarterly Bulletin of Statistics Estonia 2**

# The relationship between VAT GAP and macroeconomic indicators

**Jurgita Markevičiūtė**[1]

[1] *Vilnius University, Lithuania, jurgita.markeviciute@mif.vu.lt*

**Keywords**: VAT GAP, regression models, panel data models, disaggregation.

Each year European Union countries loses millions of Euros from the value added tax (VAT) gap. The VAT Gap is defined as the difference between the amount of VAT revenue actually collected and the theoretical amount that is expected to be collected, given the observed information on the country's economy and the actual VAT legislation [1]. The amount of VAT total theoretical liability usually is calculated using the so-called "top-down" or "bottom-up" approaches.

Since, "top-up" approaches requires the national VAT rate structure on the national accounts expenditure and investment data at the most detailed level possible or statistical data on value-added generated in each economical sector to derive expected liability, the estimate of VAT GAP is late about 2-3 years.

We focus on finding the relationship of VAT GAP with macroeconomic indicators and forecasting the VAT GAP for the current period.

For the relationship finding, we use CASE [1] VAT GAP data and various macroeconomic indicators of European Union countries. For the best model selection, we use automatic variable selection procedure based on multicollinearity checking for the regression models and we estimate panel data models for the selected sets of the variables. Our approach is similar to [2], [3] and [4].

The forecast of the VAT GAP for the current period is conducted for Lithuanian data. For this purpose, annual estimated theoretical VAT is disaggregate to quarterly data and regression models with automatic variable selection procedures are performed.

## References

[1] CASE, 2017. Study and Reports on the VAT Gap in the EU-28 Member States: 2016 Final Report. Center for Social and Economic Research, Warsaw, 18 September.

[2] Majerova, I., 2016. The Impact of Some Variables on the VAT Gap in the Member States of the European Union Company. *Oeconomia Copernicana*, 7(3), 339-355.

[3] Zidkova, H., 2014. Determinants of VAT GAP in EU. *Prague Economic Papers*, 514-530. doi:0.18267/j.pep.496

[4] Reckon, 2009. Study to Quantify and Analyze the VAT Gap in the EU-25 Member States. London: EU, Reckon.

# Semiparametric stochastic block model for longitudinal network

**Catherine Matias**[1], **Tabea Rebafka**[2] **and Fanny Villers**[3]

[1] *Centre National de la Recherche Scientifique, France, catherine.matias@math.cnrs.fr*
[2] *Sorbonne Université, France, tabea.rebafka@sorbonne-universite.fr*
[3] *Sorbonne Université, France, fanny.villers@sorbonne-universite.fr*

**Keywords**: dynamic interaction, expectation-maximization algorithm, link stream, longitudinal network, semiparametric model, variational approximation, temporal network.

We propose an extension of the stochastic block model [1] for recurrent interaction events in continuous time, where every individual belongs to a latent group and conditional interactions between two individuals follow an inhomogeneous Poisson process with intensity driven by the individuals' latent groups. We show that the model is identifiable and we estimate it with a semiparametric variational expectation-maximization algorithm [2]. We develop two versions of the method, using a nonparametric histogram approach with an adaptive choice of the partition size or kernel intensity estimators. We select the number of latent groups by an integrated classification likelihood criterion. We illustrate the utility of our approach and comment on competing methods. The full version of this work is available in [3].

## References

[1] Matias, C., Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM Proc. & Surveys* **47**, 55–74.

[2] Jordan, M., Ghahramani, Z., Jaakkola, T., Saul, L. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37**, 183–233.

[3] Matias, C., Rebafka, T., Villers, F. (to appear). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika*.

# Singular matrix-variate gamma distribution

**Tomasz Kozubowski**[1], **Stepan Mazur**[2], **Krzysztof Podgórski**[3]

[1] *University of Nevada, USA, tkozubow@unr.edu*
[2] *Örebro University, Sweden, stepan.mazur@oru.se*
[3] *Lund University, Sweden, krzysztof.podgorski@stat.lu.se*

**Keywords**: Singular matrix-variate gamma distribution, infinite divisibility, (singular) matrix-variate Laplace distribution.

We introduce a singular matrix-variate gamma distribution which is a natural extension of a singular Wishart distribution. The group properties related to infinite divisibility of the block submatrices are discussed. The continuous time stochastic process with values in the non-negative matrices is introduced. We also derive singular matrix gamma density function as well as its characteristic functions. Moreover, we introduce a (singular) matrix-variate Laplace distribution which is obtained as a product of the matrix normal distribution with the (singular) matrix-variate gamma distribution. For the extended class of matrix-variate gamma and Laplace distributions we study in full detail the infinite divisibility property both from the historical and structural perspectives.

# Network estimation in ordinary differential equation systems

**Frederik Vissing Mikkelsen**[1] **and Niels Richard Hansen**[2]

*University of Copenhagen, Denmark,* [1]*frederikrmikkelsen@gmail.com,*
[2]*Niels.R.Hansen@math.ku.dk*

**Keywords**: Chemical kinetics, inverse collocation, network inference, nonlinear least squares, ODE, systems biology.

Learning large scale nonlinear ordinary differential equation (ODE) systems from data is known to be computationally and statistically challenging. We present a framework together with the adaptive integral matching (AIM) algorithm for learning polynomial or rational ODE systems with a sparse network structure. The framework allows for time course data sampled from multiple environments representing e.g. different interventions or perturbations of the system. The algorithm AIM combines an initial penalised integral matching step with an adapted least squares step based on solving the ODE numerically. The R package episode implements AIM together with several other algorithms and is available from CRAN. It is shown that AIM achieves state-of-the-art network recovery for the *in silico* phosphoprotein abundance data from the eighth DREAM challenge with an AUROC of 0.74, and it is demonstrated via a range of numerical examples that AIM has good statistical properties while being computationally feasible even for large systems.

## References

[1] Mikkelsen, F.V. and Hansen, N.R. (2018). Learning Large Scale Ordinary Differential Equation Systems. *arXiv:1710.09308*

# Structure learning and linear stochastic differential equations

**Søren Wengel Mogensen**[1] **and Niels Richard Hansen**[2]

[1] *University of Copenhagen, Denmark, swengel@math.ku.dk*
[2] *University of Copenhagen, Denmark, niels.r.hansen@math.ku.dk*

**Keywords**: structure learning, SDEs, diffusions, causality, local independence

We consider a multivariate Ornstein-Uhlenbeck process which satisfies the stochastic differential equation

$$dX_t = AX_t dt + BdW_t,$$

where $A$ and $B$ are unknown matrices, and $B$ is diagonal. This process can be thought of as the *data-generating* process, however, it is only partially observed in the sense that we only observe a subset of the coordinate processes. The number of unobserved processes is also unknown. From the above equation, we see that each coordinate process develops as a function of the past and that in the full system the matrix $A$ encodes the *structure* of the system in the sense that for each coordinate process the placement of zeros in the corresponding matrix row determines which processes influence its development directly. However, when the system is only partially observed describing and understanding this structure becomes more involved and this work discusses how to do structure learning in this setting. If the above equation actually does describe the data-generating mechanism this will help us understand what happens in the system under interventions, that is, after fixing the paths of some of the coordinate processes.

To formalize the problem we use *local independence*, an asymmetric notion of process independence which was introduced by Schweder (1970), see [1]. In words, local independence describes how the present of one coordinate process depends on the past of other coordinate processes. Local independence has also been considered for point processes by e.g. Didelez (2008) ([2]).

We argue that the class of Ornstein-Uhlenbeck processes is tractable for such structure learning tasks as one can both explicitly calculate the conditional expectations that are needed to determine local independence and also test local independence from observational data. Given data, it is therefore possible to determine which underlying structures could have given rise to the observed local independences, accounting for the fact that only some coordinate processes are observed.

## References

[1] Tore Schweder. Composable markov processes. *Journal of Applied Probability*, 7(2):400–410, 1970.

[2] Vanessa Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society, Series B*, 70(1):245–264, 2008.

# A Bayesian meta-analysis method that corrects for publication bias

**Jonas Moss**

*University of Oslo, jonasmgj@math.uio.no*

Not all scientific studies are published. A common reason is that researchers and editors are reluctant to publish null results and results with the wrong sign. This phenomenon is captured by the term *publication bias*. For the meta-analysist, publication bias is a serious problem, as it causes systematically biased effect size estimates.

I propose a new Bayesian meta-analysis model that that explicitly models publication bias. This is done with a mixture of a truncated and an ordinary (folded) normal. The effect size of each study has been drawn from a parametric effect size distribution, which is typically non-normal and may depend on covariates. The flexibility makes it possible to study the shape of effect size distributions, something existing methods cannot do. Moreover, this allows for correction for publication bias not only for fixed effects meta-analysis, but also for mixed effects meta-analysis and meta-regression. It is also possible to estimate the degree of publication bias. I illustrate the method by estimating the true distribution of effect sizes in social psychology. The method is implemented in R and STAN, with a package forthcoming.

# Modeling global surface temperatures in terms of climate forcing and a long-memory stochastic process

**Eirik Myrvoll-Nilsen[1], Sigrunn Holbek Sørbye[2], Martin Rypdal[3], Hege-Beate Fredriksen[4], Håvard Rue[5]**

[1] *UiT - The Arctic University of Norway, eirik.myrvoll-nilsen@uit.no*
[2] *UiT - The Arctic University of Norway, sigrunn.sorbye@uit.no*
[3] *UiT - The Arctic University of Norway, martin.rypdal@uit.no*
[4] *UiT - The Arctic University of Norway, hege-beate.fredriksen@uit.no*
[5] *King Abdullah University of Science and Technology, Saudi Arabia, haavard.rue@kaust.edu.sa*

**Keywords**: Climate forcing, long memory, integrated nested Laplace approximation, Gaussian Markov random field.

Global surface temperatures are highly influenced by climate forcing variables like volcanic eruptions, solar patterns and anthropogenic effects. These variables are important, both to model and predict future temperatures. In addition, a realistic stochastic model of temperatures need to account for chaotic atmospheric dynamics. This can be achieved by using fractional Gaussian noise (fGn), which is the increment process of fractional Brownian motion. An essential property of fGn is that it exhibits long-memory properties. This imples that the resulting temperature model, incorporating climate forcing variables with fGn, becomes computationally challenging. To speed-up calculations, we provie an accurate Gaussian Markov random field approximation of fGn, using a weighted sum of just a few first-order autoregressive processes. This gives a latent Gaussian model which can be analysed efficiently using R-INLA, providing estimates of the temperature responses, the parameters of the fGn process and the contributed effects of forcing. Further, the temperature model is used to estimate the transient climate response for historical data and to give predictions of future globa surface temperature responses, given different scenarios for the climate forcing variables.

# Statistics without mathematics

**Kajsa Møllersen**

*Department of Community Medicine, Faculty of Health Sciences, UiT - The arctic university of Norway, kajsa.mollersen@uit.no*

**Keywords**: Teaching; Abstract reasoning; Concrete reasoning; Low mathematical skills

Mathematically trained students accept step-by-step logical arguments, based on assumptions or axioms. Outside mathematical and natural sciences, students more frequently reason based on perception, experience and taught knowledge. "Logic" refers to experience-based rules, not a framework for making valid conclusions. Ph.D. students of other disciplines than mathematical and natural sciences frequently use statistical methods to analyse their data. Some of them have very low mathematical skills, and are not used to abstract mathematical reasoning, which makes the challenges of teaching statistics very different.

Ph.D. students at the Health science faculty at the University of Tromsø are given a four-day mandatory course in statistics. A considerable part of the students have mathematical skills so low that they struggle with simple fractions. Increasing the mathematical skills is prevented by the limited time available, and the solution is to teach statistics without mathematics.

I choose to use two exercises that illustrate the "intuitive" (i.e. in line with experience-based reasoning) and "counter-intuitive" statistics. The first exercise is to make the students relax, boost their confidence, and give them the understanding that the statistical tools they will use are well-adapted to describe the reality. The second exercise is to make them painfully aware that experience-based reasoning will only take them so far, and that precise definitions and abstract reasoning is necessary for using statistical tools correctly.

*Exercise 1:* The students are given a handful of chocolate pieces from a large bowl, and are asked: What is your best guess of the fraction of red chocolate in the bowl?

*Exercise 2:* The students are asked to discuss/rephrase the following: "I went to my GP with a lump in my breast. I had read that the probability of cancer is 2% for women my age. The GP referred me to a specialist. But 5% of those that are referred to a specialist have cancer. Why would my GP refer me?"

These two exercises (with elaborations) are an effort to make intellectually mature students [1] prepared for abstract mathematical thinking, while avoiding the pitfall of analogy [2].

## References

[1] Miller Singley A. T., Bunge, S. A. (2014). Neurodevelopment of relational reasoning: Implications for mathematical pedagogy. *Trends in Neuroscience and Education* **3**, 33–37.

[2] Markovits, H., Doyon, C. (2011). Using analogy to improve abstract conditional reasoning in adolescents: not as easy as it looks. *European Journal of Psychology of Education* **26**, 355–372.

# Time-Dependent Transfer Entropy

**Chun-Biu Li, Joanna Tyrcha, Huong Nguyen Thu**

*Stockholm University, Sweden*

We propose a natural non-parametric generalization of transfer entropy to detect time dependent causality. This is the first time that the time-varying causal relation from one series $Y$ to another series $X$ is formulated by using the optimal history of $Y$ in addition to the one of $X$. The idea is to construct the underlying state-space network, termed $\epsilon$ machine, that extracts the relevant dynamical information of the system with minimal complexity and maximal predictivity. The power of the proposed information-theoretic approach to correctly capture the time dependent information flow will be illustrated with intuitive as well as real examples.

# A measure of dependence for multivariate stable distributions

**John P. Nolan**[1]

[1] *American University, Washington, DC, jpnolan@american.edu*

**Keywords**: multivariate stable laws, dependence.

There are several measures of dependence for stable laws, but none of them characterize independence. We discuss a measure that uses the directional scale function that does characterize independence. Properties of this measure are explored and a statistical test of independence based on sample estimations of the directional scale function is given.

# Domain selection for manifold data: Extending false discovery rates to p-values defined on continuous domains

**Niels Lundtorp Olsen**[1], **Alessia Pini**[2] **and Simone Vantini**[3]

[1] *University of Copenhagen, Denmark, niels.olsen@math.ku.dk*
[2] *Università Cattolica del Sacro Cuore, Italy, alessia.pini@unicatt.it*
[3] *Politecnico di Milano, Italy, simone.vantini@polimi.it*

**Keywords**: Functional Data Analysis, False Discovery Rate, Domain selection, Hypothesis testing, Manifold data.

A topic which is becoming more and more popular in Functional Data Analysis is local inference, i.e., the continuous statistical testing of a null hypothesis along the domain. The principal issue in this topic is the infinite amount of tested hypotheses, which can be seen as an extreme case of the multiple comparisons problem.

We define and discuss the notion of false discovery rate (FDR) in the setting of functional data defined on a compact set. Moreover, a continuous version of the Benjamini-Hochberg procedure is introduced along with a definition of adjusted $p$-value function. Some general conditions are stated, under which the continuous Benjamini-Hochberg procedure provides control of FDR.

The proposed method is applied to satellite measurements of Earth temperature. In detail, we aim at identifying the regions of the planet where temperature has significantly increased in the last decades. For comparison, we also introduce a spherical analogue of the interval-wise testing procedure (a recent non-parametric tool for local inference) able to continuously control the family-wise error rate on spherical caps. Due to the different scopes and theoretical properties of the methods, there are notable differences in the areas that are selected, with FDR being less conservative.

# Particle-based online smoothing and parameter learning in partially observed diffusion processes

**Pierre Gloaguen**[1], **Sylvain Le Corff**[2], **Jimmy Olsson**[3], **and Johan Alenlöv**[4]

[1] *IRISA, Université Bretagne-Sud, Vannes, France, pierre.gloaguen@irisa.fr*
[2] *CNRS, Université Paris-Sud, Paris, France, sylvain.lecorff@math.u-psud.fr*
[3] *KTH Royal Institute of Technology, Stockholm, Sweden, jimmyol@kth.se*
[4] *Uppsala University, johan.alenlov@it.uu.se, Uppsala, Sweden*

**Keywords**: Online learning, parameter inference, partially observed diffusion processes, sequential Monte Carlo methods, smoothing.

The *particle-based, rapid incremental smoother* (PaRIS), proposed recently in [1], approximates efficiently smoothed expectations of additive state trajectory functionals in general state-space models. The algorithm, which has typically only a linear computational complexity in the number of particles and very limited memory requirements, computes its estimates online—i.e. in a single sweep of the observed data—while avoiding completely the well-known particle-path degeneracy problem of the standard particle smoother. Thus, the sampling technique underpinning the PaRIS can be successfully used for online parameter learning via *recursive maximum likelihood estimation*; see [2]. In this talk we discuss how the PaRIS can be extended to the framework of partially observed diffusion processes via *generalised Poisson estimators*, leading to efficient online smoothing and parameter estimation in models of this sort [3].

## References

[1] Olsson, J., Westerborn, J. (2017). Efficient particle-based online smoothing in general state-space models: the PaRIS algorithm. *Bernoulli* **23**(3), 1951–1996.

[2] Olsson, J., Westerborn, J. (2017). Particle-based online estimation of tangent filters with application to parameter estimation in nonlinear state-space models. Preprint available at `https://arxiv.org/abs/1712.08466`.

[3] Gloaguen, P., Etienne, M.-P., and Le Corff, S. (2017). Efficient online sequential Monte Carlo smoother for partially observed stochastic differential equations. To appear in *EURASIP Journal on Advances in Signal Processing*.

# Empirical likelihood method for two-sample censored data

**Leonora Pahirko** [1] **and Janis Valeinis** [2]

[1] *University of Latvia, Latvia, leonora.pahirko@lu.lv*
[2] *University of Latvia, Latvia, janis.valeinis@lu.lv*

**Keywords**: right-censored data, survival analysis, empirical likelihood method.

Empirical likelihood (EL) method was introduced by Owen [1] for *iid* data and recently has been widely used to make nonparametric inference on different parameters, however it is known that EL method first was mentioned by Thomas and Grunkemeier (1975) in the context of survival analysis.

Wang and Jing [2] extended Owen's approach for randomly censored data to cover more general cases. They derived adjusted EL method for functionals of distribution function $F(\cdot)$ in the form of

$$\theta(F) = \int_0^\infty \xi(t) dF(t),$$

where $\xi$ is some measurable function which defines parameter of interest, e.g., including mean lifetime if $\xi(t) = t$.

The limiting distribution of EL ratio is a weighted $\chi^2$ distribution, where the weight depends on unknown population and needs to be estimated. This can be done using jackknife estimator of variance as proposed by Stute in 1996.

We use the plug-in EL method established by Hjort *et al.* [3]. for right-censored data for two–sample problems based on phD thesis by Valeinis [4]. Additionally we discuss some robustness properties of the newly established method.

## References

[1] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

[2] Wang, Q. H., Jing, B. Y. (2001). Empirical likelihood for a class of functionals of survival distributions with censored data. *Annals of the Institute of Statistical Mathematics* **53**, 517–527.

[3] Hjort, N. L., McKeague, I. W., Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics* **37**, 1079–1111.

[4] Valeinis, J. (2007). *Confidence bands for structural relationship models.* Phd thesis, Goerg-August-University of Goettingen.

# Exact long term descriptions of diffusions under Markovian switching

**Abhishek Pal Majumder**[1] **and Filip Lindskog**[2]

[1] *Stockholm University, Sweden, majumder@math.su.se*
[2] *Stockholm University, Sweden, lindskog@math.su.se*

**Keywords**: Regime switching diffusion process, Ornstein-Uhlenbeck process, Stochastic recurrence equation.

Regime switching stochastic processes are indispensable in modelling several financial problems [3],[1], [2] when the underlying model parameters are observed to fluctuate over time. In diffusion contexts we study one such example: Ornstein-Uhlenbeck process when drift and diffusion coefficients switch throuh Markovian regimes denoted by process $X_\cdot$ through functional forms respectively denoted by $a(\cdot), b(\cdot)$. The exact long time asymptotics are established in all possible cases (positive recurrent, null recurrent and transient) that are determined by the sign of the expected drift $E_\pi a(\cdot)$ (respectively at $E_\pi a(\cdot) > 0, = 0, < 0$) under stationarity of the underlying regime process $X_\cdot$. Previously only different types of tail behaviours were investigated under stable regime $E_\pi a(\cdot) > 0$ while here we get the explicit characterization of the stationary distribution. Additionally time limit results for integrals of form $I_t := \int_0^t b^2(X_s) e^{-2\int_s^t a(X_{s_1})ds_1} ds$ are given in three different regimes along with several other examples of diffusions. This work connects with several results of affine invariant solution of stochastic recurrence equation of type $X \stackrel{d}{=} AX + B$ [4] that can be interpreted as the stationary solution of a random coefficient Autoregressive process.

## References

[1] Andrew V Carter , Benjamin Hansen and Douglas G Steigerwald. Detecting regime switching: Analytic power. *Technical Report, Department of Economics, Columbia University*, 2016.

[2] Niloufar Abourashchi, Iain Clacher, Mark C Freeman, David Hillier, Malcolm Kemp and Qi Zhang. Pension plan solvency and extreme market movements: a regime switching approach *The European Journal of Finance*, 22(13): 1292 – 1319, 2016. Taylor & Francis

[3] Holger Fink, Yulia Klimova, Claudia Czado and Jakob Stöber. Regime switching vine copula models for global equity and volatility indices. *Econometrics*, 5(1): 3–17, 2017. Multidisciplinary Digital Publishing Institute

[4] Dariusz Buraczewski, Ewa Damek and Thomas Mikosch. Stochastic models with power-law tails. Springer Publication, 2016.

# Optimal Estimation in Mixed Membership Stochastic Block Models

**Evgeny Marshakov[1], Nikita Mokrov[2], Maxim Panov[1] and Roman Ushakov[2]**

[1] *Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia*
[2] *Institute for Information Transmission Problems of RAS, Moscow, Russia*
*m.panov@skoltech.ru*

**Keywords**: Random graphs, overlapping community detection, stochastic block model.

Community detection is an important problem in modern network analysis. It has wide applications in analysis of social and biological networks, designing network protocols and many other areas. Recently, much attention has been paid to detection of overlapping communities, where each node in a network may belong to multiple communities. We note that the majority of overlapping community detection methods come with no guarantees on their performance.

This paper considers the parameter estimation problem in Mixed Membership Stochastic Block Model (MMSB) [1], which is a quite general instance of random graph model allowing for overlapping community structure. We consider algorithm *successive projection overlapping clustering* (SPOC) which combines the ideas of spectral clustering and geometric approach for separable non-negative matrix factorization in MMSB. The algorithm starts from the spectral embedding based on the adjacency matrix of the graph, then finds nearly pure nodes via successive projection algorithm (SPA) [2] and finally reconstructs community memberships via least-squares fit. SPOC is provably consistent under MMSB with general conditions on the parameters of the model [3] and it is also shown to perform well experimentally in comparison to other algorithms. In this work, we provide lower bounds on the rates of parameter estimation in MMSB and develop an improved version of SPOC algorithm achieving minimax optimal rates.

**Acknowledgment** The reported study was funded by RFBR according to the research project # 18-37-00489.

## References

[1] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. *Mixed membership stochastic blockmodels*. Journal of Machine Learning Research, 9(Sep):1981 – 2014, 2008.

[2] Mario Araujo et al. *The successive projections algorithm for variable selection in spectroscopic multicomponent analysis*. Chemometrics and Intelligent Laboratory Systems, 57(2):65 – 73, 2001.

[3] Maxim Panov, Konstantin Slavnov and Roman Ushakov. *Consistent Estimation of Mixed Memberships with Successive Projections*, "Complex Networks & Their Applications VI", Springer, 2017.

# Graphical posterior predictive classifier: Bayesian model averaging using sequential Monte Carlo methods

**Tatjana Pavlenko**[1] **and Felix Leopoldo Rios**[2]

[1] *KTH Royal Institute of Technology, Sweden, pavlenko@math.kth.se*
[2] *KTH Royal Institute of Technology, Sweden, flrios@math.kth.se*

**Keywords**: Graphical models, Bayesian inference, predictive classification, decomposable graphs

In this talk, we present a multi-class graphical Bayesian predictive classifier of [3] that incorporates the uncertainty in the model selection into the standard Bayesian formalism. For each class, the dependence structure underlying the observed features is represented by a set of decomposable Gaussian graphical models. Emphasis is then placed on the Bayesian model averaging which takes full account of the class-specific model uncertainty by averaging over the posterior graph model probabilities. Even though the decomposability assumption severely reduces the model space, the size of the class of decomposable models is still immense, rendering the explicit Bayesian averaging over all the models infeasible. To address this issue, we consider the particle Gibbs strategy proposed in [2] for posterior sampling from decomposable graphical models which utilizes the Christmas tree algorithm of [1] as proposal kernel. We also derive the a strong hyper Markov law which we call the hyper normal Wishart law that allows to perform the resultant Bayesian calculations locally. The proposed predictive graphical classifier reveals superior performance compared to the ordinary Bayesian predictive rule that does not account for the model uncertainty, as well as to a number of out-of-the-box classifiers.

## References

[1] Olsson J., Pavlenko T. and Rios F. L. (2017). Generating junction trees of decomopsable graphs with the Christmas tree algorithm. DiVA: diva2:1141466.

[2] Olsson J., Pavlenko T. and Rios F. L. (2017). Bayesian structure learning in graphical models using sequential Monte Carlo. DiVA: diva2:1141468.

[3] Pavlenko T. and Rios F. L. (2017). Graphical posterior predictive classifier: Bayesian model averaging with particle Gibbs. arXiv:1707.06792v2.

# Flipping the classroom

**Inger Persson**

*Uppsala University, Sweden, inger.persson@statistik.uu.se*

It is well known that a higher degree of student activity leads to better results, increased attendance and engagement. Nevertheless, many statistics instructors are still giving traditional lectures. In this talk I'll present the Flipped classroom pedagogy, as one very hands-on alternative to increase student activity. I'll show how I've chosen to use the technique in a multivariate analysis course, and what kinds of results that has lead to, sharing my experiences and most useful practical tips.

# Event study testing with cross-sectional correlation due to partially overlapping event windows

**Seppo Pynnönen[1], James W. Kolari[2] and Bernd Pape[3]**

[1] *University of Vaasa, Finland, sjp@uwasa.fi*
[2] *Texas A&M University, USA, JKolari@mays.tamu.edu*
[3] *University of Vaasa, Finland, bepa@uwasa.fi*

**Keywords**: abnormal returns, clustering robust standard errors, cross-sectional correlation, clustered event days.

This article examines the issue of cross-sectional correlation due to partially overlapping event windows. Kolari and Pynnönen (2010) find that in the case of event-date clustering, even relatively low cross-sectional correlation among abnormal returns seriously biases standard tests to over-reject the null hypothesis of zero average abnormal return. While in Kolari and Pynnönen the underlying assumption is that the firms share the same event window in calendar time, we generalize the approach to a case where the windows overlap in calendar time only partially. It turns out that this generalization implies a modification to their test statistics that accounts for the average percentage of the overlaps in the event windows. The paper proposes also estimators of the average correlation and covariance-variance ratios needed for the adjustment in this more general case of partial overlaps.

## References

[1] Kolari, J.W. and Pynnönen, S. (2010). Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* **23**, 3996–4025.

# Modelling ultra-high frequency trade durations with the ACD model

**Markus Belfrage**[1] **and Gunnar Rosenqvist**[2]

[1] *Hanken School of Economics, Finland, markus.belfrage@gmail.com*
[2] *Hanken School of Economics, Finland, gunnar.rosenqvist@hanken.fi*

**Keywords**: Autoregressive Conditional Duration model, Finance, High frequency trading, Model selection, Market microstructure.

Engle and Russell [2] introduced the autoregressive conditional duration (ACD) model for modelling time series of durations. The conditional expected duration is modelled as a function of past duration and past expected duration, similar to how the GARCH (generalized autoregressive conditional heteroscedasticity) model models volatility, and indeed the ACD model for durations and GARCH model for volatilities resemble and share properties with each other. This gave rise to a fast growing literature in which various developments of the ACD model were introduced and applied to financial data, and corresponding tests were also developed. For reviews see e.g. Pacurar [4] and Hautsch [3]. Belfrage [1] provides an R package for a number of these models and tests. The last decade has witnessed dramatic changes of the financial trading landscape with among other things ultra-high frequency of trading and enormous sets of trading data. We will study a data set comprising all trades in Apple (AAPL) on NASDAQ on one day (October 4, 2016, n = 21 247). We will fit a range of different ACD type models to this set of high frequency trade durations, compare the fit of the models with each other, and strive to identify to what degree the different models are able to accommodate various aspects of the data. We analyse the data with a precision of nano ($10^{-9}$) seconds as well as with a precision of hundredths ($10^{-2}$) of a second. We conclude that ACD type models have obvious difficulties in reproducing the shortest durations and argue that durations below approximately a hundredth of a second are dominated by high frequency algorithmic trading and need other types of models.

## References

[1] Belfrage, M. (2016). *ACDm: Tools for Autoregressive Conditional Duration Models*. R package version 1.0.5.

[2] Engle, R. F., and J. R. Russell (1998). Autoregressive Conditional Duration: A New Approach for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127-1162.

[3] Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Springer Science & Business Media.

[4] Pacurar, M. (2008). Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys* **22**, 711-751.

# Causal interpretation in survival analysis

**Kjetil Røysland**

*University of Oslo, Norway, kjetil.roysland@medisin.uio.no*

It has recently been emphasized, see (Hernan 2010), that the common interpretation of hazards in survival analysis, as risk of death during an interval $[t, t + \Delta)$ for an individual alive at $t$, is often not true. Hazards have a built-in bias since when conditioning on recent survival, we actually condition on a collider that is likely to open a non-causal pathway from the exposure to the event of interest. It does not even matter if the underlying model is causal. A simple example in (Aalen, Cook, Røysland 2015) shows that this can also be a problem in RCTs. There are potentially dramatic consequences from this because modern survival analysis has very much been centered around the concept of hazards. As matter of fact, clinicians have been advised to focus on hazards, instead of other measures, since they where thought to contain more dynamical information.

In order to to deal with this situation, we suggest a flexible method to estimate other parameters than the hazard from causal models in survival analysis. These parameters do not have the built the built-in bias as the hazards have, and will often allow interpretations that are more intuitive to clinicians. Our method combines the non-parametric additive hazard model with stability theorems for stochastic differential equations to define a class of estimators that are easily implemented on a computer. We also present a proof that these estimators are consistent, given fairly weak assumptions.

# Bayesian quantile regression for discrete observations

**Håvard Rue**[1]

[1] *King Abdullah University of Science and Technology, Saudi Arabia, haavard.rue@kaust.edu.sa*

**Keywords**: Bayesian quantile-regression, asymmetric Laplace distribution

Quantile regression, i.e. modeling conditional quantiles of some covariates and other effects through the linear predictor, has typically been carried out exploiting the asymmetric Laplace distribution (ALD) as a working "likelihood". In the Bayesian framework, this is highly questionable as the posterior variance is affected by the artificial ALD "likelihood". With continuous responses, we can reparameterize the likelihood in terms of a $\alpha$-quantile, and let the $\alpha$-quantile depend on the linear predictor. We can then do model based quantile regression with little effort using the `R-INLA` package (`www.r-inla.org`) doing approximate Bayesian inference for latent Gaussian models, and trust the quantile regression posterior in the same way as when doing parametric mean regression.

For discrete variables, like Poisson and (negative) Binomial, there is no continuous relationship between quantiles and distribution's parameters, hence model based quantile regression seems no longer possible. In this talk I will discuss how to resolve this issue, so that we can do model based quantile regression also for discrete responses. I will present some examples that also demonstrate how the parametric approach almost resolves the quantile crossing problem.

This is joint work with Tullia Padellini, Sapienza University of Rome, Italy.

# On some stochastic neuronal models

**Adeline Leclercq Samson**

*Universite Grenoble Alpes, adeline.leclercq-samson@univ-grenoble-alpes.fr*

**Keywords**: Neuronal data, Hawkes processes, Diffusions, Statistical learning

Neuronal data are of different types: at the scale of a single neuron (intra-cellular data), of a set of hundred neurons (extra-cellular data) or at the scale of a region of the brain (fMRI, EEG, MEG data). Neuronal data are intrinsically stochastic.

In this lecture, I will present an overview of stochastic models that have been proposed to describe neuronal data at different scales: stochastic differential equations either elliptic or hypoelliptic, Hawkes processes, functional processes.

I will explain the link between the different models, and recent advances in understanding networks of neurons, how different populations of neurons interact and synchronize. Then I will give an overview of statistical inference methods that have been proposed in the literature to estimate these models from the available data.

# Gaussian approximations of SDEs in Metropolis adjusted Langevin algorithms

**Simo Särkkä**[1]

[1] *Aalto University, simo.sarkka@aalto.fi*

**Keywords**: Metropolis adjusted Langevin algorithm, Gaussian assumed density approximation, Markov chain Monte Carlo, Metropolis–Hastings.

Metropolis adjusted Langevin algorithm (MALA) [1] is a Markov chain Monte Carlo (MCMC) method that can be used to generate samples from otherwise intractable probability distributions arising, for example, in Bayesian statistics. MALA is an Metropolis–Hastings algorithm, where the proposal density is selected to be the transition density of a stochastic differential equation (SDE) of the form

$$d\theta = f(\theta)\, dt + dW.$$

Above, $\theta(t) \in \mathbb{R}^d$, $W$ is a $d$-dimensional Wiener process and $f(\theta) = \frac{1}{2}\nabla \log \pi(\theta)$, where $\pi$ is the (unnormalized) target density, typically known only up to a normalization constant. The stationary distribution of the SDE has the density $p(\theta) \propto \pi(\theta)$ and thus by simulating samples from the SDE over sufficiently long intervals we can generate samples from the target distribution. Because the transition density of the SDE can rarely be evaluated in closed form, in MALA, the SDE is typically approximated with the Euler–Maruyama method [1]. However, Euler–Maruyama is inherently a local approximation and thus only works for short intervals.

In this talk, we study a class of methods where, instead the Euler–Maruyama, the SDE is approximated using Gaussian assumed density approximations often used in stochastic filtering context [2]. In these approximations, the SDE is approximated with a Gaussian process, whose mean $m(t)$ and covariance $P(t)$ evolve according to the differential equations [2]

$$dm/dt = \hat{\mathrm{E}}_\theta[f(\theta)], \qquad dP/dt = \hat{\mathrm{E}}_\theta[f(\theta)\,(\theta - m)^T] + \hat{\mathrm{E}}_\theta[(\theta - m)\, f^T(\theta)] + I,$$

where $\hat{\mathrm{E}}_\theta$ denotes the expectation with respect to a Gaussian distribution with mean $m$ and covariance $P$. As these approximations can also be used to form approximations to the transition densities of SDEs, they provide the means to construct Metropolis–Hastings algorithms which use the approximate transition densities as proposal densities. We show how to construct practical MCMC methods from these constructions, evaluate the methods numerically, and discuss some theoretical properties of the algorithms.

## References

[1] Roberts, G. O., Rosenthal J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **60**(1), 255–268.

[2] Särkkä, S., Sarmavuori, J. (2013). Gaussian filtering and smoothing for continuous-discrete dynamic systems. *Signal Processing* **93**(2), 500–510.

# Exact inference on Bayesian networks including gamma distributed variables

**Ivar Simonsson** [1] **and Petter Mostad**[2]

[1] *Chalmers University of Technology and Gothenburg University, Sweden, simonssi@chalmers.se*
[2] *Chalmers University of Technology and Gothenburg University, Sweden, mostad@chalmers.se*

**Keywords**: Bayesian networks, exact inference, variable elimination

Exact inference on Bayesian networks (BNs) can be performed using the variable elimination algorithm. This algorithm performs a series of operations, including analytical marginalization, on different components of the network and in principal it can be done on all kinds of graphical networks. However, in order to run the algorithm in practice, one has to impose rather strict conditions on the structure and the variables of the network. The algorithm can easily be made to work for the base case in which all the variables of the BN have finite scope. It turns out that it is also doable on what is called *conditional linear Gaussian BNs*, see [1], [2]. These are networks that can include both variables with finite scope and Gaussian variables (with restrictions on the connections between them, for details see [2]).

In [3] one step towards extending the applicability of the algorithm was taken by making it possible to include a gamma distributed variable so that the precision of the Gaussian variables also could be modeled. A serious limitation in that paper was that only one gamma variable could be included and that this variable had to model the precision of *all* Gaussian variables in the network.

In this work we present one way to move forward so that more freedom in modeling the precision of Gaussian variables can be allowed. Multiple gamma variables can now be included to model the precision of different Gaussian variables. One problem when doing this is that analytical marginalizations is no longer possible in all stages of the algorithm and one has to rely on numerical integration. However, we show that one can order the variable elimination, using some additional tricks, so that the dimensions of these numerical integrations are manageable.

## References

[1] Cowell, R. G., Dawid, P., Lauritzen, S. L., Spiegelhalter, D. J. (2006). Probabilistic networks and expert systems: Exact computational methods for Bayesian networks. Springer Science & Business Media.

[2] Koller, D., Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT press.

[3] Simonsson, I., Mostad, P. (2016). Exact Inference on Conditional Linear Γ-Gaussian Bayesian Networks. *Conference on Probabilistic Graphical Models* (pp. 474-486).

# Prediction of spatial functional random processes: Comparing functional and spatio-temporal kriging approaches

**Johan Strandberg**[1], **Sara Sjöstedt de Luna**[2] **and Jorge Mateu**[3]

[1] *Umeå University, Sweden, johan.strandberg@umu.se*
[2] *Umeå University, Sweden, sara.sjostedt.de.luna@umu.se*
[3] *Universitat Jaume I, Spain, mateu@mat.uji.es*

**Keywords**: Functional kriging, prediction, spatial functional random processes, spatio-temporal kriging.

We present and compare functional and spatio-temporal (Sp.T.) kriging approaches to predict spatial functional random processes (which can also be viewed as Sp.T. random processes). Comparisons with respect to computational time and prediction performance via functional cross-validation is evaluated, mainly through a simulation study but also on a real data set. We restrict comparisons to Sp.T. kriging versus ordinary kriging for functional data (OKFD), since the more flexible functional kriging approaches, pointwise functional kriging (PWFK) and the functional kriging total model, coincide with OKFD in several situations. In particular, we formulate conditions under which we show that OKFD and PWFK coincide.

From the simulation study, it is concluded that the prediction performance for the two kriging approaches in general is rather equal for stationary Sp.T. processes. However, functional kriging tends to perform better for small sample sizes, while Sp.T. kriging works better for large sizes. For non-stationary Sp.T. processes, with a common deterministic time trend and/or time varying variances and dependence structure, OKFD performs better than Sp.T. kriging irrespective of the sample size. For all simulated cases, the computational time for OKFD was considerably lower compared to those for the Sp.T. kriging methods.

# Probabilistic Approaches To Statistics of Manifold Data

**Stefan Sommer**[1]

[1] *Department of Computer Science, University of Copenhagen, sommer@di.ku.dk*

**Keywords**: statistics on manifolds, stochastic processes

An alternative to performing statistical inference in manifolds by optimizing least squares criterions such as those defining the Fréchet mean is to optimize the likelihood of data. This approach emphasizes maximum likelihood means over Fréchet means, and it in general allows generalization of Euclidean statistical procedures defined via the data likelihood. While parametric families of probability distributions are generally hard to construct in nonlinear spaces, transition densities of stochastic processes provide a geometrically natural way of defining data likelihoods. Examples of this includes the Riemannian Brownian motion and anisotropic generalizations [3], both leading to generalizations of the Euclidean normal distribution, and stochastic coadjoint motion [1] when Lie group structure is present. In the talk, we discuss likelihood based inference on manifolds and procedures for approximating data likelihood by simulation of manifold and Lie group valued diffusion bridges [2].

## References

[1] Alexis Arnaudon, Darryl D. Holm, and Stefan Sommer, *A Geometric Framework for Stochastic Shape Analysis*, accepted for Foundations of Computational Mathematics, arXiv:1703.09971 [cs, math] (2018).

[2] Stefan Sommer, Alexis Arnaudon, Line Kuhnel, and Sarang Joshi, *Bridge Simulation and Metric Estimation on Landmark Manifolds*, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, Lecture Notes in Computer Science, Springer, September 2017, pp. 79–91 (en).

[3] Stefan Sommer and Anne Marie Svane, *Modelling anisotropic covariance using stochastic development and sub-Riemannian frame bundle geometry*, Journal of Geometric Mechanics **9** (2017), no. 3, 391–410 (en).

# Careful prior specification avoids incautious inference for log-Gaussian Cox point processes

**Sigrunn Holbek Sørbye[1], Janine Illian[2], Daniel Simpson[3], David Burslem[4] and Håvard Rue[5]**

[1]*UiT The Arctic University of Norway, Norway, sigrunn.sorbye@uit.no*
[2]*University of St.Andrews, UK, jbi@st-andrews.ac.uk*
[3]*University of Toronto, Canada, dp.simpson@gmail.com*
[4]*University of Aberdeen, UK, d.burslem@abdn.ac.uk*
[5]*King Abdullah University of Science and Technology, Saudi Arabia, haavard.rue@kaust.edu.sa*

**Keywords**: Hyperprior specification, penalised complexity prior, `R-INLA`, spatial point process modelling

Hyperprior specifications for random fields in spatial point process modelling can have a major impact on the statistical inference and the conclusions made. We consider model fitting of a discretized log-Gaussian Cox process to spatial point patterns, relative to spatial covariate data. From an ecological point of view, an important aim of the analysis is to assess significant associations between the covariates and the point pattern intensity of a given species.

To improve control and facilitate meaningful interpretations of the results, we consider a reparameterised model, combining a spatially structured field and an unstructured random field as one component. This component has one hyperparameter accounting for marginal variance, while an additional hyperparameter governs the fraction of variance explained by the spatially structured term. The hyperparameters are assigned penalised complexity priors, which can be tuned intuitively by user-defined scaling parameters. Also, proper scaling of the spatially structured field makes the analysis invariant to grid resolution. The given modelling approach provides a clear interpretation of the hyperparameters and significance of covariates can be studied for a range of hyperprior specifications. We illustrate the approach using `R-INLA`, analysing covariate effects on point patterns formed by rainforest tree species.

# Maximum likelihood estimation for stochastic differential equations with random effects

**Michael Sørensen**[1]

[1] *University of Copenhagen, Denmark, michael@math.ku.dk*

**Keywords**: Bayesian inference, diffusion bridge simulation, discretely sampled diffusions, likelihood inference, mixed effects.

Simulation-based methods for maximum likelihood estimation for stochastic differential equations with random effects are presented. The focus is on a Gibbs sampler that uses the new simple method for simulation of diffusion bridges developed in [1] and [2]. The Gibbs sampler is illustrated by an application to an Ornstein-Uhlenbeck process with random effects in the drift and in the diffusion coefficient. An EM-algorithm is briefly considered.

The bridge simulation method is explained in detail. By a novel application of classical coupling methods, approximate, but often very accurate, diffusion bridges can be simulated. These approximate bridges are used as proposal for an easily implementable MCMC algorithm that produces exact diffusion bridges. An advantage of the new bridge simulation method, compared to the methods of Roberts and coworkers, is that it works well for diffusion bridges in long time intervals. This is because the computational complexity of the method is linear in the length of the interval. In the statistical application this property implies that the method also works well at low sampling frequencies.

The lecture is based on joint work with Mogens Bladt, Fernando Baltazar Larios and Samuel Finch.

## References

[1] Bladt, M. and Sørensen, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, **20**, 645–675.

[2] Bladt, M., Finch, S. and Sørensen, M. (2016). Simulation of multivariate diffusion bridges. *J. Roy. Statist. Soc., Ser. B*, **78**, 343 – 369.

# Competition in randomly growing processes

**Alexandre Stauffer**

*University of Bath, United Kingdom, a.stauffer@bath.ac.uk*

We consider random growth processes that compete for space over time. This is by now a classical topic in probability theory. The usual situation is that when the two processes have different speeds of growth, then one of the processes wins against the other. It is quite rare to find natural models where both processes coexist forever. In this talk I will discuss a random growth model, which we introduced as a tool to studying a famous model of dendritic growth from physics. This growth model can also be regarded as a model for blocking the spread of fake news in a network. We will discuss the behavior of this processes, its phase transition and the occurrence of coexistence. This is based on joint works with Elisabetta Candellero (Warwick) and Vladas Sidoravicius (NYU Shanghai).

# Pension Saving Decision Making under Lifetime and Investment Uncertainty

**Mogens Steffensen**

*University of Copenhagen, Denmark, mogens@math.ku.dk*

We discuss the design of pension products. The theoretical context is decision making under uncertainty with notions of preferences like risk aversion, habit formation, and resolution of uncertainty. The practical questions we address concern guaranteed benefits, smoothing of investment returns and...would you actually like to know when you are going to die?

# On application of weighted Kolmogorov-Smirnov statistics to the problems of signal detection and classification in sparse models.

**Natalia Stepanova[1], Tatjana Pavlenko[2], and Lee Thompson[3]**

[1] *Carleton University, Canada, nstep@math.carleton.ca*
[2] *KTH Royal Institute of Technology, Sweden, pavlenko@math.kth.se*
[3] *Carleton University, Canada, LeeThompson@cmail.carleton.ca*

**Keywords**: goodness-of-fit tests, weighted empirical processes, binary classification, feature selection, sparse models.

In this talk, we show how goodness-of-fit test statistics based on sup-functionals of weighted empirical processes can be effectively applied to various problems of signal detection and classification in sparse models. The weight functions employed are Chibisov-O'Reilly functions, as proposed in [1] and [3], and Erdős-Feller-Kolmogorov-Petrovski upper-class functions of a Brownian bridge, as studied in [2]. The obtained results demonstrate the advantage of our approach to studying the problems of signal detection and classification in sparse models over a common approach that utilizes regularly varying weight functions.

## References

[1] Chibisov, D. M. (1964). Some theorems on the limiting behaviour of empirical distribution functions. *Selected Transl. Math. Statist. Probab.* **6**, 147–156.

[2] Csörgő, M., Csörgő, S., Horváth, L., & Mason, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14**, 31–85.

[3] O'Reilly, N. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probab.* **2**, 642–651.

# Models and inference for on-off data via clipped Ornstein-Uhlenbeck processes

**Emil Aas Stoltenberg**[1] **and Nils Lid Hjort**[2]

*University of Oslo, Norway,* [1]*emilas@math.uio.no,* [2]*nils@math.uio.no*

**Keywords**: On-off data, Ornstein-Uhlenbeck, clipped processes, quasi likelihood, composite likelihood, event history.

Most people are in good health most of the time and, unfortunately, sick some of the time. In this paper we study a model where for a sample of individuals whose health statuses are governed by independent latent Ornstein-Uhlenbeck process. A person is sick if the process is above a certain threshold, and in good health otherwise, and it is only this only "clipped" zero-one version of the process that is actually observed. Moreover, the sample is not continuously monitored, and the health status of an individual is only ascertained at certain points in time. These time points might be different for each of the individuals, they need not be equidistant, or they might be generated according to a stochastic process (independent of the underlying OU-processes).

A clipped Gaussian process is no longer Markov [3] and likelihood inference is not straightforward. We propose and study a version of the quasi-likelihood [2] for inference on the parameters governing the underlying OU-processes. Large-sample properties of the estimators are presented, and the methods are applied to a data set of 926 Brazilian children followed up over a period of 15 months. The results of this analysis are compared and contrasted with the counting process approach, using versions of Aalen's additive risk model, of [1] to the same data set.

## References

[1] Borgan, Ø., Fiaccone, R.L., Henderson, R., Barreto, M.L. (2007). Dynamic Analysis of Recurrent Event Data with Missing Observations, with Application to Infant Diarrhoea in Brazil. *Scandinavian Journal of Statistics* **34**, 53–69.

[2] Hjort, N.L., Varin C. (2008). ML, PL, QL in Markov Chain Models. *Scandinavian Journal of Statistics* **35**, 64–82.

[3] Slud, E. (1989). Clipped Gaussian Processes Are Never M-step Markov. *Journal of Multivariate Analysis* **29**, 1–14.

# On the modified Palm version

**Hermann Thorisson**

*University of Iceland, Iceland, hermann@hi.is*

**Keywords**: Palm theory, shift-coupling, mass-stationarity.

The Palm version of a stationary random measure is an important tool in probability. It is however not well known that there are in fact two Palm versions, with related but different interpretations. For lack of better terms, call the well known version standard and the less known version modified. In this talk we shall focus on the modified Palm version and its interpretation. The concepts of shift-coupling and mass-stationarity will play a key role.

# Csörgő-Csörgő-Horwáth-Mason based inference for high-dimensional linear models

**Annika Tillander** [1]

[1] *Linköping University, Sweden, annika.tillander@liu.se*

**Keywords**: Goodness-of-fit, Higher Criticism, Mixture detection, Multiple comparisons

In the high-dimensional setting it is challenging to identify the informative features especially when sparse and weak. Given some measure of information strength a threshold is required. Let $X_1, X_2, \ldots$ be real-valued iid random variables with a continuous cumulative distribution function (CDF) $F$. The empirical distribution function (EDF) based on $X_1, \ldots, X_n$ is denoted $\mathbb{F}_n(t) = n^{-1}\sum_{i=1}^n \mathbb{I}(X_i \leq t)$. We consider testing the hypothesis of goodness-of-fit

$$H_0 : F(t) = F_0(t) \forall t \in \mathbb{R} \text{ vs. } H_1 : F(t) \neq F_0(t) \text{ for some } t \in \mathbb{R}$$

The following test statistic suggested by [1]

$$T_n(q) = \sup_{0<F_0(t)<1} \frac{\sqrt{n}|\mathbb{F}_n(t) - F_0(t)|}{q\left(F_0(t)\right)}$$

The probability integral transformation for $T_n(q)$, given that $H_0$ is true, is then for each $n$ equal in distribution to $\sup_{0<u<1} \frac{\sqrt{n}|\mathbb{U}_n(u)-u|}{q(u)}$ where $\mathbb{U}_n(u) = n^{-1}\sum_{i=1}^n \mathbb{I}(U_i \leq u$ and $q(u) = \sqrt{u(1-u)\log\log\left(1/(u(1-u))\right)}$ with $U_1, \ldots, U_n$ being iid uniform $U(0,1)$[3]. This is compared to a type of standard deviation proportional weight function suggested by [2]

$$\mathrm{HC}_n = \sup_{0<u<\alpha_0} \frac{\sqrt{n}\left(\mathbb{U}_n(u) - u\right)}{\sqrt{u(1-u)}}, 0 < \alpha_0 < 1$$

## References

[1] Csörgő, M., Csörgő, S., Horwáth L., Mason D.M. (1986). Weighted Empirical and Quantile Processes *The Annals of Probability* **14:1** 31–85.

[2] Donoho, D., Jin J. (2004). Higher Criticism for Detecting Sparse Heterogeneous Mixtures *The Annals of Statistics* **32:3** 962–994.

[3] Stepanova, N., Pavlenko, T. (2014). Goodness-of-fit tests based on sup-functionals of weighted empirical processes. *ArXiv e-prints* 1406.0526.

# Empirical likelihood for some robust methods

**Janis Valeinis**[1]

[1] *University of Latvia, Latvia, janis.valeinis@lu.lv*

**Keywords**: Empirical likelihood, robustness, trimmed mean

In this work we introduce and discuss empirical likelihood (EL) for some robust methods including M- and L-estimates. Firstly we establish some new robust two-sample procedures based on smooth Huber estimators and trimmed means. Additionally we mention some extension to ANOVA-type procedures [1]. Secondly we discuss some new results regarding EL for linearly or smoothly trimmed mean introduced already by [2]. The advantages of using such robust estimators are shown by extensive simulation study.

## References

[1] Velina, M., Valeinis, J., Greco, L. and Luta, G. (2016). Empirical Likelihood-based ANOVA for trimmed means. *International Journal of Environmental Research and Public Health* **13**(10), 953.

[2] Stigler, S.M. (1973). The asymptotic distribution of the trimmed mean. *The Annals of statistics* **1**(3), 472–477.

# Monotonicity for multi-range percolation on oriented trees

**Daniel Valesin**

*University of Groningen, The Netherlands, D.Rodrigues.Valesin@rug.nl*

Percolation theory contains a variety of monotonicity-related conjectures which are easy and natural to state but challenging to prove. In this talk, we will survey a few of these problems. We will then specialize to the setting of Bernoulli percolation on multi-range oriented trees. The trees we consider are oriented regular trees where besides the usual short bonds, all bonds of a certain length are added. Independently, short bonds are open with probability $p$ and long bonds are open with probability $q$. We study properties of the critical curve which delimits the set of pairs $(p, q)$ for which there is percolation. We also show that this curve decreases with respect to the length of the long bonds. Joint work with Bernardo N.B. de Lima and Leonardo T. Rolla.

# On Bayesian uncertainty quantification

**Aad van der Vaart**[1]

[1] *Universiteit Leiden, Netherlands, avdvaart@math.leidenuniv.nl*

**Keywords**: Credible set, confidence set, nonparametric estimation, Bayesian adaptation, sparsity

We give an overview of theoretical results that justify, or not, the use of a posterior distribution of a high- or infinite-dimensional parameter as a method for uncertainty quantification. The posterior distribution is the conditional distribution of the parameter given the data when the parameter is thought of as having been generated by a prior. For the non-Bayesian it is just a random distribution over the parameter space, whose spread could be used to form the equivalent of a confidence set. We start by noting that in the nonparametric situation this is only justified if the prior does not oversmooth the true parameter. In practice, the smoothness of the prior is chosen dependent on the data so as to adapt the procedure to unknown smoothness or sparsity level. We show that then uncertainty quantification is correct for a large set of, but not all, true parameters. This is related to the incompatibility of adaptation and uniform coverage by any statistical procedure, including Bayesian ones. We illustrate the results by the examples of Gaussian process priors for curve or surface estimation and inverse problems, Dirichlet mixtures for density estimation, and the spike-and-slab and horseshoe priors for the estimation of high-dimensional sparse models, every of which we briefly introduce.

# Structural Change Detection in Multivariate Systems and the Recent Financial Crisis Period

**Lauri Viitasaari**

[1] *University of Helsinki, Finland, lauri.viitasaari@iki.fi*

**Keywords**: Structural change detection, multivariate regression, nonparametric, financial crisis

Structural change detection problems are often encountered in analytics and econometrics, where the performance of a model can be significantly affected by unforeseen changes in the underlying relationships. Although these problems have a comparatively long history in statistics, the number of studies done in the context of multivariate data under nonparametric setting is still small. In this talk we propose a consistent method for detecting multiple structural changes in a system of related regressions over a d-dimensional variable space. As an example we consider change point detection in the context of the recent financial crisis period.

# Mixed linear models with latent variables

**Dietrich von Rosen**[1,2]

[1] *Swedish University of Agricultural Sciences, Sweden, dietrich.von.rosen@slu.se*
[2] *Linköping University, Sweden*

**Keywords**: Bilinear regression model, random effects, rank restrictions.

Bilinear models with three types of effects are considered: fixed effects, random effects and latent variable effects. Explicit estimators are proposed.

# Bilinear regression with rank restrictions on the mean and the dispersion matrix

**Tatjana von Rosen**[1] **and Dietrich von Rosen**[2]

[1] *Stockholm University, Sweden, tatjana.vonrosen@stat.su.se*
[2] *Swedish University of Agricultural Sciences; Linköping University, Sweden, dietrich.von.rosen@slu.se*

**Keywords**: growth curve model, likelihood based estimates, rank restrictions, singular dispersion matrix.

In this work, we consider a general bilinear regression model with rank restrictions imposed on the mean-parameter matrix and on the dispersion matrix. Likelihood based estimates for the mean parameters and covariance matrix in this type of models are obtained. A numerical example is provided to illustrate the obtained results.

## References

[1] Potthoff, R.F., Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[2] von Rosen, T., von Rosen, D. (2017). On estimation in some reduced rank extended growth curve models. Mathematical Methods of Statistics, **26**, 299–310.

[3] Reinsel, G.C., Velu, R.P. (1998). *Multivariate Reduced-Rank Regression. Theory and Applications.* Lect. Notes Stat., 136, Springer, New York.

# On model fitting and estimation of stationary processes

**Marko Voutilainen**

*Aalto University, Finland, marko.voutilainen@aalto.fi*

**Keywords**: Stationary process, ARMA process, ARCH process, model fitting, parameter estimation.

Stationary processes form an important class of stochastic processes that has been extensively studied in the literature. Their applications include modelling and forecasting numerous real life phenomenon including natural disasters, sustainable energy sources, sales and market movements.

One of the most essential families of stationary processes is the ARMA family. When modelling existing data with an ARMA process, the first step is to fix the orders of the model. After that, one can estimate the related parameters by using standard methods such as maximum likelihood (ML) or least squares (LS) estimators. The final step is to conduct various diagnostic tests in order to determine the quality of the model. In this talk we present a novel way of fitting a model to a data that is assumed to be a realization from a discrete time stationary process. Our approach is based on a recently proved AR(1) characterisation of strictly stationary processes, where the noise is not assumed to be white. As a result, we obtain a closed form consistent estimator of the model parameter and its asymptotic normality under general conditions. In comparison to conventional ARMA modeling, in general, the ML and LS estimators do not admit closed form representations.

The method is also applicable in estimation of the ARCH model with weighted liquidity. We obtain consistent estimators of the model parameters under some assumptions that are mostly natural.

# Model misspecification and bias for inverse probability weighting and doubly robust estimators

**Ingeborg Waernbaum**[1] **and Laura Pazzagli**[2]

[1] *UmeåUniversity, Sweden, ingeborg.waernbaum@umu.se*
[2] *Karolinska Institute, Sweden, laura.pazzagli@ki.se*

**Keywords**: Average causal effects, comparing biases, propensity score, robustness.

In the causal inference literature a class of semi-parametric estimators is called robust if the estimator has desirable properties under the assumption that at least one of the working models is correctly specified. A standard example is a doubly robust estimator that specifies parametric models both for the propensity score and the out- come regression. As an alternative, we propose a crude analytical approach to study the large sample bias of estimators when all models are assumed to be approximations of the true data generating process, i.e., all models are misspecified. We apply our approach to three prototypical estimators, two inverse probability weighting (IPW) estimators, using a misspecified propensity score model, and a doubly robust (DR) estimator, using misspecified models for the outcome regression and the propensity score. To analyze the question of when using two misspecified models are better than one we derive necessary and sufficient conditions for when the DR estimator has a smaller bias than the simple IPW estimator and when it has a smaller bias than the IPW estimator with stabilized weights. For the conditions in the comparisons the covariance between the propensity score model error and the conditional outcomes plays an important role. For example, for the DR estimator to improve upon the stabilized IPW estimator a necessary condition is that the outcome is misspecified within an interval defined by the true conditional out- come and the absolute value of the covariance. The results are illustrated in a simulation study.

# Focused model selection and inference using robust estimators with a nonparametric alternative

**Sam-Erik Walker**[1] **and Nils Lid Hjort**[1]

[1]*Department of Mathematics, University of Oslo, Norway,*
*samerikw@math.uio.no, nils@math.uio.no*

**Keywords**: Asymptotic theory, focused information criterion, maximum weighted likelihood estimator, model selection, nonparametric alternative, robust statistics.

Selection of parametric models based on general information criteria like Akaike's information criterion (AIC), Bayesian information criterion (BIC), and similar, is a well-established practice within the statistical science. Over the years, more specific model selection criteria have also been developed, like the so-called focused information criterion, or FIC, where models are selected based on specifying a focus parameter – a certain parameter or function of parameters deemed most important in a given statistical setting. Up to now FIC has been mostly based on maximum likelihood estimator methodology; see [1], [3]. Such procedures are efficient under model conditions, but not robust. In this work we extend the theory and application of the FIC to the use of robust estimators, such as the density power divergence estimator, and to a newly developed maximum weighted likelihood estimator, cf. [2]. By comparing with nonparametric alternatives, using FIC, we may perform focused model selection and inference in situations where the model might be misspecified, or where data might be contaminated with atypical values or outliers. A real data example is provided to illustrate the new methodology.

## References

[1] Claeskens, G., Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.

[2] Hjort, N.L., Walker, S.-E. (2018). Estimation and model selection via weighted likelihoods. Manuscript, Department of Mathematics, University of Oslo.

[3] Jullum, M., Hjort, N.L. (2017). Parametric or nonparametric: The FIC approach. *Statistica Sinica* **27**, 951–981.

# Simulation of Prostate Cancer Prevalence

**Marcus Westerberg**

*Uppsala University, Sweden, marcus.westerberg@math.uu.se*

**Keywords**: Simulation, prostate cancer, screening, prevalence, incidence, multiple imputation

Early detection of prostate cancer (PCa) in asymptomatic men using prostate-specific antigen (PSA) screening has been shown to drastically increase the incidence of low risk PCa while decreasing the incidence of metastatic disease (mPCa). The uptake of PSA testing in Sweden has been substantial, with great variation between counties, despite the absence of national recommendations for PSA-based PCa screening.

The aim of this study is to model the effect of opportunistic PSA screening on prostate cancer incidence, prevalence and mortality in Sweden, by risk stage, in order to produce a prognosis for 2030 under two different screening scenarios: fixed opportunistic screening as of 2012 and virtually no screening as of 1992.

Data on men with PCa diagnosed between 1970-2012 was obtained through record linkage using the Swedish personal numer from the Swedish Cancer Register, the National Prostate Cancer Register (NPCR) and other high quality health care registers. In addition, data on men without PCa was obtained from controls associated with each man with PCa, and demographic data was obtained from Statistics Sweden (SCB) including a prognosis for 2013-2030.

The PCa risk stage was categorized into five risk categories and the method of multiple imputation using chained equations (MICE) was used in case of missing data.

The simulation model consisted of three parts, PCa incidence, mortality among men with PCa and mortality among men without PCa diagnosis. Each generation, defined by birth year and county of residence, was followed from the age of 40 to 100. Prostate cancer incidence was modelled using four piece-wise exponential models, where the incidence of low risk PCa per numbers at risk was assumed to reflect the amount of PSA screening. The incidence of the higher risk categories was modelled using observed low and intermediate incidence history per numbers at risk. Estimation was performed using penalized Poisson regression with age and age-dependent screening history 0-20 years back as covariates. Similarly, mortality of men with PCa diagnosis was modelled on individual level by a multinomial model, and mortality among men without PCa was modelled using logistic model and estimated on data on controls, censoring for PCa diagnosis.

Preliminary results suggest that the effect of opportunistic PSA screening compared to no screening on incidence of mPCa is weak and that opportunistic screening will lead to a considerably higher prevalence of low risk PCa. Thus, current screening practice is suboptimal and provides only a small benefit compared to no screening.

# Prediction of high current for thermography in maintenance of electrical networks

**Per Westerlund[1], Jesper Rydén[2], Patrik Hilber[1], Tommie Lindquist[3]**

[1] *KTH Royal Institute of Technology, Sweden, perw@kth.se*
[2] *Uppsala University, Sweden*
[3] *Svenska kraftnät (Swedish National Grid), Sweden*

**Keywords**: autocorrelation function, electrical substations, thermography.

A standard method to check the condition of an electrical substation (where overhead lines meet) is thermography. It captures the infrared radiation from heated points. Then it is possible to detect contacts in bad condition. The accuracy of thermography decreases with current [2, Paper VII].

Thus it is necessary to find times when the current is as high as possible. [1] has shown a model to predict high currents. If the maintenance engineer decides to do thermography at a substation, he/she can use the model to plan in which week, which weekday and when to send out a technician.

If the current is too low during the thermography, when should it redone? Should the technician be sent there again the next day, the next week or the next month? Then there is need to estimate the autocorrelation function.

An analysis of the autocorrelation function of the current during one year at the same point as in [1] in the Swedish electrical transmission grid shows there is variation depending on the time of the day and on the day of the week. The correlation also decreases with the separation in weeks from the measurement date.

After four weeks, the correlation is low, between -0.2 and 0.3. The correlation varies that much in a 24-hour period. It takes 6 hours for the correlation to decrease by two thirds. An analysis of another point with more data (10 years) shows a similar behaviour, although the correlation only reaches -0.1.

Thus if the current is too low for the thermography to be reliable, the technician should come back after a month in order to have current that is not strongly correlated to the previous current. However, [1] shows that the probability of having a high current is low for the studied point in a Swedish substation. So even if the correlation decreases with time, the current will probably be too low the next time.

## References

[1] Westerlund, P., Hilber, P. and Tommie Lindquist. Prediction of current in a substation in order to schedule thermography. In *PMAPS – Probability Methods Applied to Power Systems*, Beijing, China, October 2016.

[2] Lindquist, T. On reliability and maintenance modelling of ageing equipment in electric power systems. Doctoral thesis, KTH Royal Institute of Technology, 2008.

# Resolving the Lord's Paradox with Predictive and Causal Arguments

**Priyantha Wijayatunga**

*Umeå University, Sweden, priyantha.wijayatunga@umu.se*

**Keywords**: averages, effects, confounding, regression models, causal diagrams.

We give an explanations to the Lord's paradox [2] using both the ordinary least square regression models and the causal graphical models. We argue that it is not a paradox at all, if the regression parameters are interpreted as predictive effects in respective models. The regression parameters can also be interpreted as causal effects with stricter conditions. Sometimes paradoxical conclusions arise when we are not clear about the purpose of asking the question that is answered through statistical models. Our regression based solution is by showing how to derive a super-model from a given or estimated sub-model, whenever its residuals can be modelled with other potential predictors. Our causal arguments that oppose to those in the current literature [1] comply with Lord's initial arguments [3]. We also discuss the difficulties of deriving causal diagrams for Lord's data and some of the other similar data contexts. We argue that dependencies inherited in data are more important than independencies when answering causal questions.

## References

[1] Pearl, J. (2016). Lord's Paradox Revisted - (Oh Lord Kumbaya!). *Journal of Causal Inference* **4(2)**. DOI: 10.1515/jci-2016-0021

[2] Wijayatunga, P. (2017a). Resolving the Lord's Paradox. In: Grzegorczyk, M. and Ceoldo, G. (ed), *Proceedings of the 32nd International Workshop on Statistical Modelling* **II**, 223–226.

[3] Wijayatunga, P. (2017b). Discussion on the Paper Titled *Beyond objective and subjective in statistics* by Gelman and Hennig" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **180(4)**, 1023–1024.

# Non-penalized variable selection in high-dimensional settings via generalized fiducial inference

**Jonathan P. Williams**[1] **and Jan Hannig**[2]

[1] *University of North Carolina at Chapel Hill, USA, jpwill@live.unc.edu*
[2] *University of North Carolina at Chapel Hill, USA, jan.hannig@unc.edu*

Standard penalized methods of variable selection and parameter estimation rely on the magnitude of coefficient estimates to decide which variables to include in the final model. However, coefficient estimates are unreliable when the design matrix is collinear. To overcome this challenge an entirely new perspective on variable selection is presented within a generalized fiducial inference framework. This new procedure is able to effectively account for linear dependencies among subsets of covariates in a high-dimensional setting where $p$ can grow almost exponentially in $n$, as well as in the classical setting where $p \leq n$. It is shown that the procedure very naturally assigns small probabilities to subsets of covariates which include redundancies by way of explicit $L_0$ minimization. Furthermore, with a typical sparsity assumption, it is shown that the proposed method is consistent in the sense that the probability of the true sparse subset of covariates converges in probability to 1 as $n \to \infty$, or as $n \to \infty$ and $p \to \infty$. Very reasonable conditions are needed, and little restriction is placed on the class of possible subsets of covariates to achieve this consistency result.

In addition to establishing the variable selection methodology in the high-dimensional linear model setting, in a forthcoming second paper we modify the procedure for use in the high-dimensional vector autoregressive time-series setting. Preliminary findings on its asymptotic properties are presented.

# CLT for the intrinsic Fréchet mean on compact Riemannian manifolds

**Andrew T.A. Wood**

*University of Nottingham, UK, Andrew.Wood@nottingham.ac.uk*

**Keywords**: central limit theorem; cut locus; density estimation.

We will begin with a brief review of intrinsic and extrinsic Fréchet means on manifolds. Relevant asymptotic results for such means include the central limit theorem (CLT) results given in [1] and [2]. However, the intriguing, non-standard results on the circle given in [3], which will be summarized briefly in the talk, demonstrate that the CLT results in [1] and [2] do not give a complete picture of what happens in the case of intrinsic Fréchet means. The remainder of the talk will focus on a new CLT for intrinsic Fréchet means, which agrees with the results of [3] on the circle and also applies to a general class of compact Riemannian manifolds. Estimation of the asymptotic covariance matrix in non-standard cases will also be considered. This talk is based on joint work with Thomas Hotz (Ilmenau) and Huiling Le (Nottingham).

Note: new results concerning CLTs on manifolds in a complementary direction are given in [4].

## References

[1] Bhattacharya, R.N. and Patrangenaru, V. (2003). Large sample theory of instrinsic and extrinsic sample means on a manifold, I. *Annals of Statistics*, **31**, 1-29.

[2] Bhattacharya, R.N. and Patrangenaru, V. (2005). Large sample theory of instrinsic and extrinsic sample means on a manifold, II. *Annals of Statistics*, **33**, 1225-1259.

[3] Hotz, T. and Huckemann, S. (2015). Intrinsic Means on the Circle: Uniqueness, Locus and Asymptotics. The Annals of the Institute of Statistical Mathematics, **67**, 177-193

[4] Eltzner, B., Huckemann, S. F. (2018). A Smeary Central Limit Theorem for Manifolds with Application to High Dimensional Spheres. *arXiv 1801.06581*. Submitted for publication.

# Teaching statistics using technology

**Lena Zetterqvist**

*Lund University, Sweden, lena.zetterqvist@matstat.lu.se*

In this talk I will describe how we use technology in basic courses in probability and statistics for engineering students at LTH. Some examples are digital exercises and tests, video clips with theory or solutions to exercises, digital tutoring for computer exercises with Matlab and projects. In most of our courses the students work with a mixture of digital, paper-and-pen and computer exercises. In one course the students bring their own laptops to all exercise sessions using the programming language Matlab. I will also shortly refer to the literature where potential benefits when using technology are identified and where the role of technology in enhancing the learning experience is discussed.

# On Generalized Multivariate Linnik Distribution

**Igor V. Zolotukhin**[1]

[1] *Shirshov Institute of Oceanology, Russian Academy of Sciences, Russia,*
*Igor.Zolotukhin@gmail.com*

**Keywords**: Linnik distribution, Marshall-Olkin multivariate exponential distribution, strictly stable distributions, mixture, characteristic function.

The classical Linnik distribution (see [1]) is one of the basic symmetric distributions given by the characteristic function

$$\varphi(\theta) = \frac{1}{1 + \eta\,|\theta|^{\alpha}}$$

where $\eta > 0$ is a scale parameter, $0 < \alpha \leq 2$ is a stability parameter. The case of $\alpha = 2$ sets a Laplace distribution.

Are known several multivariate versions of this distribution, for example, see [2]. The following multivariate generalization of Linnik's law is proposed.

Let the vector $X = (X_1, \ldots, X_n)$ consist of several independent random variables $X_i$ with strictly stable distributions with parameters $\alpha_i, \eta_i, \beta_i$, and the vector $Z = (Z_1, \ldots, Z_n)$ is an independent from $X$ random vector having the Marshall-Olkin multivariate exponential distribution.

Let us introduce the notation:

$$A = (\alpha_1^{-1}, \ldots, \alpha_n^{-1}), \quad Z^A = (Z_1^{\alpha_1^{-1}}, \ldots, Z_n^{\alpha_n^{-1}}).$$

Then distribution of the vector $V = Z^A X$ is called **Generalized Multivariate Linnik Distribution**.

The following properties of the distribution were derived:

- known multidimensional distributions (in particular, the Anderson distribution) are the special cases of the proposed one;
- the explicit expression for the characteristic function has been found;
- the recursive formulas to calculate the characteristic functions for the projections of the distribution on any coordinate hyperplane also has been found;
- it has been shown that the distribution can be recursively restored by distributions of their univariate components.

The research was performed in the framework of the state assignment of FASO Russia (theme No. 0149-2018-0014).

## References

[1] Linnik, Yu. V. (1963). Linear forms and statistical criteria, I, II. *Selected Translations in Math. Statist. Probab.* **3**, 1–90.

[2] Anderson, D. N. (1992). A multivariate Linnik distribution. *Statistics and probability letters* **14(4)**, 333–336.

# On generalized total least squares estimators for big errors-in-variables models

**Silvelyn Zwanzig**

*Uppsala University, zwanzig@math.uu.se*

**Keywords**: Big data, measurement errors, total least squares, covariance matrix estimation.

We consider a high dimensional linear model, where the number of replications $n$ is less than the number $p$ of covariates. Furthermore the covariates are observed with additional error. Under normal distributed errors the total least squares estimator is the maximum likelihood estimator. In case of $n << p$ the total least squares estimator and the naive least squares estimator coincide. They deliver an overfit of the model. Thus alternative estimators are needed.

Assume a data set

$$Z_i = (W_{1,i}, \ldots, W_{p,i}, y_i), \;\; i = 1, \ldots, n$$

generated from the structural relation

$$y_i = \sum_{j=1}^{p} \beta_j^T X_{i,j} + \varepsilon_i, \;\; W_{i,j} = X_{i,j} + \delta_{i,j}$$

where $X_{i,j}$ are latent and i.i.d. r.v. with expectation zero and variance 1, the errors are independent of each other have expectation zero and variance $\sigma^2$. Then the observations $Z_i$ are i.i.d $(p+1)$–dimensional random vectors with expectation zero and

$$Cov(Z) = \Sigma(\beta) = L_\beta L_\beta^T + \sigma^2 I_{p+1}.$$

The matrix $L_\beta L_\beta^T$ has eigenvalues $0, 1, \ldots, 1, 1 + \|\beta\|^2$ , the eigenvector belonging to the eigenvalue zero is $(\beta^T, -1)^T$. Thus the estimated eigenvector belonging to the minimal eigenvalue of $\Sigma(\beta)$ delivers an estimator of the parameter $\beta$ in the structural relation. In case of $n > p$ we can apply the sample covariance and we obtain the total least squares estimator. For big data the sample covariance is singular. In the talk generalizations of the total least squares estimator are proposed basing on alternative covariance matrix estimators. A similar approach was proposed using rank estimates under $n > p$ in [1].

## References

[1] S. Zwanzig. On a consistent rank estimate in a linear structural model. *Tatra Mt. Math. Publ.* **51**, 191–202.

# Index