

Conference on Conditional Independence Structures and Extremes

October 12 - 14, 2016

Garching, Germany



Programme and Abstracts

Detailed Programme

Wednesday, October 12, 2016

08:30 - 09:00	<i>Registration</i>
09:00 - 09:15	Welcome Words
09:15 - 10:00	Steffen Lauritzen: Total positivity and Markov structures
10:00 - 10:45	Peter Green: Bayesian inference with and about decomposable models
10:45 - 11:15	<i>Coffee Break</i>
11:15 - 12:00	John Einmahl: A continuous updating weighted least squares estimator of tail dependence in high dimensions
12:00 - 12:45	Johan Segers: Regularly varying Markov trees
12:45 - 14:00	<i>Lunch Break</i>
14:00 - 14:45	Claudia Klüppelberg: Max-linear models on graphs
14:45 - 15:30	Nadine Gissibl: Estimation of max-linear models on directed acyclic graphs
15:30 - 16:00	<i>Coffee Break</i>
16:00 - 16:45	Bernd Schölkopf: Toward causal machine learning
16:45 - 17:30	Jonas Peters: Invariances and causality
17:30 - 20:00	<i>Poster Session with drinks and snacks</i> (Location: Faculty of Mathematics, ground floor, Room 00.10.011)

Thursday, October 13, 2016

09:00 - 09:45	Adrien Hitz: Graphical modelling of extremes
09:45 - 10:30	Sven Buhl: Semiparametric estimation for max-stable space-time processes
10:30 - 11:00	<i>Coffee Break</i>
11:00 - 11:45	Bernd Sturmfels: Convexity in tree spaces
11:45 - 12:30	Caroline Uhler: Learning DAGs based on sparsest permutations
12:30 - 14:00	<i>Lunch Break</i>
14:00 - 14:45	Sebastian Engelke: Extremes on River Networks
14:45 - 15:30	Holger Rootzén: Multivariate GP distributions: portfolio risk estimation, prediction of flu epidemics, spatial rainfall modelling
15:30 - 16:00	<i>Coffee Break</i>
16:00 - 16:45	Nanny Wermuth: On palindromic Ising models with graph structure
16:45 - 17:30	Daniel Straub: Conditional independence in risk analysis of engineered systems
18:00 - open end	<i>Conference Dinner</i>

Friday, October 14, 2016

09:00 - 09:45	Piotr Zwirnik: The correlation space of Gaussian latent tree models and model selection without fitting
09:45 - 10:30	Alberto Roverato: Path weights, networked partial correlations and their application to the analysis of genetic interactions
10:30 - 11:00	<i>Coffee Break</i>
11:00 - 11:45	Philippe Naveau: Detecting changes in climate records
11:45 - 12:30	Richard Davis: Applications of distance correlation to time series
12:30 - 14:00	<i>Lunch</i>

Abstracts of Talks

Sven Buhl, Technical University of Munich, Germany

SEMIPARAMETRIC ESTIMATION FOR MAX-STABLE SPACE-TIME PROCESSES

Abstract: We propose a semiparametric estimation procedure based on a closed form expression of the extremogram (cf. [2], [3], [4]) to estimate the model parameters in a max-stable space-time process. We establish the asymptotic properties of the resulting parameter estimates. A simulation study shows that the proposed procedure works well for moderate sample sizes. Finally, we apply this estimation procedure to fitting a max-stable model to radar rainfall measurements in a region in Florida. This modeling procedure helps to quantify the extremal properties of the space-time observations.

This talk is based on a joint work with Richard Davis, Claudia Klüppelberg and Christina Steinkohl.

References:

- [1] S. Buhl, R. A. Davis, C. Klüppelberg and C. Steinkohl (2016). Semiparametric estimation for max-stable space-time processes. *In preparation*.
 - [2] S. Buhl and C. Klüppelberg (2016). Limit theory for the empirical extremogram of random fields. *In preparation*.
 - [3] Y. Cho, R. A. Davis and S. Ghosh (2016). Asymptotic Properties of the Empirical Spatial Extremogram. *Scandinavian Journal of Statistics*.
 - [4] R. A. Davis and T. Mikosch (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15**(4):977-1009.
-

Richard A. Davis, Columbia University, New York, USA

APPLICATIONS OF DISTANCE CORRELATION TO TIME SERIES

Abstract: The use of empirical characteristic functions for inference problems, including estimation in some special parametric settings and testing for goodness of fit, has a long history dating back to the 70s (see for example, Feuerverger and Mureika (1977), and Csörgő (1981)). More recently, there has been renewed interest in using empirical characteristic functions in other inference settings. The distance covariance and correlation, developed by Szekely and Rizzo (2007) for measuring dependence and testing independence between two random vectors, are perhaps the best known illustrations of this. We apply these ideas to stationary univariate and multivariate time series to measure lagged auto- and cross-dependence in a time series. Assuming strong mixing, we establish the relevant asymptotic theory for the empirical auto- and cross-distance correlation functions. We also apply the auto-distance correlation function (adcf) to the residuals of an autoregressive processes as a test of goodness of fit. Under the null that an autoregressive model is true, the limit distribution of the empirical adcf can differ markedly from the corresponding one for an iid sequence. We illustrate the use of the empirical adcf for testing dependence and cross-dependence of time series in a variety of different contexts.

This talk is based on joint work with Muneya Matsui, Thomas Mikosch, and Phyllis Wan.

John H.J. Einmahl, Tilburg University, The Netherlands

A CONTINUOUS UPDATING WEIGHTED LEAST SQUARES ESTIMATOR OF TAIL DEPENDENCE IN HIGH DIMENSIONS

Abstract: Likelihood-based procedures are a common way to estimate tail dependence parameters. They are not applicable, however, in non-differentiable models such as those arising from recent max-linear structural equation models. Moreover, they can be hard to compute in higher dimensions. An adaptive weighted least-squares procedure matching nonparametric estimates of the stable tail dependence function with the corresponding values of a parametrically specified proposal yields a novel minimum-distance estimator. The estimator is easy to calculate and applies to a wide range of sampling schemes and tail dependence models. In large samples, it is asymptotically normal with an explicit and estimable covariance matrix. The minimum distance obtained forms the basis of a goodness-of-fit statistic whose asymptotic distribution is chi-square. Extensive Monte Carlo simulations confirm the excellent finite-sample performance of the estimator and demonstrate that it is a strong competitor to currently available methods. The estimator is then applied to disentangle sources of tail dependence in European stock markets.

This talk is based on a joint work with Anna Kiriliouk and Johan Segers.

Sebastian Engelke, Ecole Polytechnique Fédérale de Lausanne, Switzerland

EXTREMES ON RIVER NETWORKS

Abstract: Max-stable processes are suitable models for extreme events that exhibit spatial dependencies. The dependence measure is usually a function of Euclidean distance between two locations. In this talk, we model extreme river discharges on a river network in the upper Danube catchment, where flooding regularly causes huge damage. Dependence is more complex in this case as it goes along the river flow. For non-extreme data a Gaussian moving average model on stream networks was proposed by Ver Hoef and Peterson (2010). Inspired by their work, we introduce a max-stable process on the river network that allows flexible modeling of flood events and that enables risk assessment even at locations without a gauging station. Recent methods from extreme value statistics are used to fit this process to a big data set from the Danube area.

Nadine Gissibl, Technical University of Munich, Germany

ESTIMATION OF MAX-LINEAR MODELS ON DIRECTED ACYCLIC GRAPHS

Abstract: We consider structural equation models where all random variables can be written as a max-linear function of their parents and independent noise variables. For the corresponding graph we assume that it is a directed acyclic graph. We develop estimation methods for this model class. Firstly, we assume that the structure of the graph is given. We derive a generalized maximum likelihood estimator for the weights of the max-linear structural equation model. Secondly, we address the important question of identifiability and estimation of its causal structure. We propose a statistical method and an algorithm based on our theoretical findings.

References:

- [1] Gissibl, N. and Klüppelberg, C. (2015). *Max-linear models on directed acyclic graphs*. URL <http://arxiv.org/abs/1512.07522>.
- [2] Gissibl, N., Klüppelberg, C. and Lauritzen, S. (2016). *Estimation of max-linear models on directed acyclic graphs*. In preparation.
- [3] Gissibl, N., Klüppelberg, C. and Otto, M. (2016). Tail dependence of regularly varying models on directed acyclic graphs. In preparation.

Peter Green, UTS, Sydney, Australia and University of Bristol, UK.

BAYESIAN INFERENCE WITH AND ABOUT DECOMPOSABLE MODELS

Abstract: The advantages of being able to assume decomposability in the manipulation of graphical models, and classical inference about them, are well known. This talk explores some of the implications of decomposability for Bayesian structural inference about graphical models for both continuous and discrete data, focussing particularly on (conjugate) graph priors and posterior sampling.

This talk is based on joint work with Alun Thomas.

References:

- [1] S. Byrne and A. P. Dawid (2015). Structural Markov graph laws for Bayesian model uncertainty. *The Annals of Statistics*, **43**:1647–1681.
 - [2] A. P. Dawid and S. L. Lauritzen (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, **21**:1272–1317.
 - [3] A. M. Fitch, M. B. Jones and H. Massam (2014). The performance of covariance selection methods that consider decomposable models only. *Bayesian Analysis*, **9**:659–684.
 - [4] P. J. Green and A. Thomas (2013). Sampling decomposable graphs using a Markov chain on junction trees. *Biometrika*, **100**:91–110.
-

Adrien Hitz, University of Oxford, UK

GRAPHICAL MODELLING OF EXTREMES

Abstract: The problem of inferring the distribution of a random vector given that its norm is large requires modelling a homogeneous limiting density. In this talk, I will present an approach based on graphical models which is suitable for high-dimensional vectors. I will introduce the notion of asymptotic conditional independence and relate it to a factorisation of the limiting density, generalising Hammersley-Clifford theorem.

This talk is based on a joint work with Prof. Robin Evans.

References:

- [1] A. Hitz and R. Evans (2015). One-Component Regular Variation and Graphical Modeling of Extremes. URL <http://arxiv.org/abs/1506.03402>.
-

Claudia Klüppelberg, Technical University of Munich, Germany

MAX-LINEAR MODELS ON GRAPHS

Abstract: We consider a new structural equation model, where all random variables can be written as a max-linear function of their parents and independent noise terms. We assume that the dependence structure of the corresponding vector can be modeled by a directed acyclic graph. We show that the multivariate distribution is max-linear and characterize all max-linear models, which are generated by a structural equation model. We detail the relation between the coefficients of the structural equation model and the max-linear coefficients. Furthermore, we derive minimal representations of our model and investigate certain subgraphs arising from order relations between (or equality of) different components of the vector.

This talk is based on joint work with Nadine Gissibl.

References:

- [1] Gissibl, N. and Klüppelberg, C. (2015) Max-linear models on directed acyclic graphs. URL <http://arxiv.org/abs/1512.07522>.
-

Steffen Lauritzen, University of Copenhagen, Denmark

TOTAL POSITIVITY AND MARKOV STRUCTURES

Abstract: Positive associations between random variables can be described in a variety of ways. One of strongest is multivariate total positivity of order two (MTP2) introduced by Karlin and Rinott [2]. The MTP2 property is stable under a number of operations, including marginalization and conditioning. The lecture investigates how the property interacts with conditional independence structures, implying, for example, that any strictly positive MTP2 distribution becomes faithful to its pairwise independence graph. In addition we shall study how this property is manifested in Gaussian, discrete, and conditional Gaussian distributions, and give a number of examples of such distributions.

The lecture is based on joint work with S. Fallat, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik; [2].

References:

- [1] S. Karlin and Y. Rinott (1980). Classes of orderings of measures and related correlation inequalities I: multivariate totally positive distributions. *Journal of Multivariate Analysis* **10**: 467–498.
- [2] S. Fallat, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik (2016). Total positivity in Markov structures. URL <http://arxiv.org/abs/1510.01290>. To appear in *The Annals of Statistics*.

Philippe Naveau, CNRS IPSL LSCE, France

DETECTING CHANGES IN CLIMATE RECORDS

Abstract: In this talk, I focus primarily on detecting changes in temperature and precipitation over long time periods, either from instrumental or climate model records. Here, the term extremes refers to record events in annual maxima of daily maxima time series. Records are often the most important events in terms of impact and choosing records as the event of interest can improve the inference of the ratio of two small probability events. The gain is particularly substantial if we leverage the tools and hypothesis used in Extreme Value Theory, especially from records properties in the context of max-stable distributions.

To detect changes in climate records, we assume that we have long observational time series and different climate models at our disposal, and any of those climate models has been run under two different scenarios, say natural and all forcings respectively. To illustrate our approach, the Fraction of Attributable Risk, an often used indicator in event attribution studies, is modified and tailored to handle records. We discuss the advantages of our method through theoretical results, simulation studies, observational time series and outputs from numerical climate models. If time allowed, comparisons with other non-parametric techniques will be presented.

This talk is based on a joint work with Francis Zwiers, Aurelien Ribes, Alexis Hannart and Pascal Yiou.

References:

- [1] Hannart, A., J. Pearl, Otto, F. E. L., Naveau, P. and Ghil, M (2015). Causal counterfactual theory for the attribution of weather and climate-related events. *BAMS*, in press.
- [2] Hegerl, G. and Zwiers F. (2011). Use of models in detection and attribution of climate change. *Wiley interdisciplinary reviews: climate change* 2(4): 570–591. DOI 10.1002/wcc.121.
- [3] Naveau P. Zwiers F., Ribes A. (2011). Revising return periods for record events in a climate event attribution context. *Submitted*.
- [4] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. New York, NY: Springer.

Jonas Peters, MPI Tübingen, Germany

INVARIANCES AND CAUSALITY

Abstract: In causal inference, we are interested in the system's behavior under an intervention: what happens to gene B, for example, if we delete gene A? If gene A is causal for gene B, then predicting B from A works equally well in the intervened and in the non-intervened scenario; this does not hold if gene B is causal for gene A, for example.

In our talk, we discuss two implications of this well-known observation:

- (i) Causal models are more stable against a change of environments and may be used for problems in domain adaptation.
- (ii) If we are given data from different environments, we can look for such invariant models and draw conclusions about the underlying causal structure.

Holger Rootzén, Chalmers University of Technology, Gothenburg, Sweden

MULTIVARIATE GP DISTRIBUTIONS: PORTFOLIO RISK ESTIMATION, PREDICTION OF FLU EPIDEMICS, SPATIAL RAINFALL MODELLING

Abstract: This talk surveys work in progress on new parametric multivariate generalized Pareto models for extreme episodes. These models, perhaps surprisingly, have simpler and more tractable likelihoods than standard max-stable models, and permit use of the entire maximum likelihood machinery for estimation, testing, and model checking. I will show how the models can contribute to three areas. The first one is fast and consistent risk estimation aimed at aiding financial portfolio selection. Influenza is a main cause of death, and a major cause of stress on the health care system. The second contribution is models and prediction methods for the development of flu epidemics. Finally, I will discuss how GP distributions can provide spatial models for extreme environmental events.

The talk is based on a joint work with Anna Kiriliouk, Maud Thomas, Johan Segers, and Jennifer Wadsworth.

Alberto Roverato, University of Bologna, Italy

PATH WEIGHTS, NETWORKED PARTIAL CORRELATIONS AND THEIR APPLICATION TO
THE ANALYSIS OF GENETIC INTERACTIONS

Abstract: Genetic interactions confer robustness on cells in response to genetic perturbations. This often occurs through molecular buffering mechanisms that can be predicted using, among other features, the degree of coexpression between genes, commonly estimated through marginal measures of association such as Pearson or Spearman correlation coefficients. However, marginal correlations are sensitive to indirect effects and often partial correlations are used instead. Yet, partial correlations convey no information about the (linear) influence of the coexpressed genes on the entire multivariate system, which may be crucial to discriminate functional associations from genetic interactions. To address these two shortcomings, here we propose to use the edge weight derived from the covariance decomposition over the paths of the associated gene network. We call this new quantity the networked partial correlation and use it to analyze genetic interactions in yeast. More concretely, in its well-characterized leucine biosynthesis pathway and on a previously published data set of genome-wide quantitative genetic interaction profiles. In both cases, networked partial correlations substantially improve the identification of genetic interactions over classical coexpression measures.

This talk is based on a joint work with Robert Castelo, University Pompeu Fabra, Spain.

Bernhard Schölkopf, MPI Tübingen, Germany

TOWARD CAUSAL MACHINE LEARNING

Abstract: In machine learning, we use data to automatically find dependences in the world, with the goal of predicting future observations. Most machine learning methods build on statistics, but one can also try to go beyond this, assaying causal structures underlying statistical dependences. Can such causal knowledge help prediction in machine learning tasks? We argue that this is indeed the case, due to the fact that causal models are more robust to changes that occur in real world datasets. We touch upon the implications of causal models for machine learning tasks such as semi-supervised learning, domain adaptation, and transfer learning.

We also present an application to the removal of systematic errors for the purpose of exoplanet detection. Machine learning currently mainly focuses on relatively well-studied statistical methods. Some of the causal problems are conceptually harder, however, the causal point of view can provide additional insights that have substantial potential for data analysis.

Johan Segers, Catholic University of Louvain, Belgium

REGULARLY VARYING MARKOV TREES

Abstract: Extreme values of regularly varying Markov chains can be described in terms of the limiting conditional distribution of the normalized chain given that it is large at a particular time instant [1]. The limit distributions are called forward and backward tail chains, according to the time direction considered [2]. Viewing a chain as a tree consisting of a single, long branch, we seek for generalizations to general Markov trees, i.e., random vectors whose dependence structure is governed by a tree representing a set of conditional independence relations together with a collection of bivariate distributions along the tree edges [3]. As for Markov chains, we find that extremal dependence of such Markov trees can be described in terms of a collection of tail trees, each tree describing the limit distribution of the rescaled Markov tree given that its value at a particular node is large. Moreover, the time-change formula for tail chains generalizes to a relation between these tail trees. Tail trees can be used to compute quantities such as the number of nodes in the graph affected by a shock at a particular node or the probability that a particular part of the graph will be affected by a shock in another part of the graph. Moreover, specifying the graph structure and the bivariate distributions along the edges provides a construction method for max-stable models.

This talk is based on joint work with Gildas Mazo.

References:

- [1] Basrak, B. and Segers, J. (2009) Regularly varying multivariate time series. *Stochastic Processes and Their Applications*, **119**:1055–1080.
- [2] Janssen, A. and Segers, J. (2014) Markov tail chains. *Journal of Applied Probability*, **51**(4):1133–1153.
- [3] S. L. Lauritzen (1996). *Graphical Models*. Oxford University Press, Oxford.

Daniel Straub, Engineering Risk Analysis Group, Technical University of Munich, Germany

CONDITIONAL INDEPENDENCE IN RISK ANALYSIS OF ENGINEERED SYSTEMS

Abstract: Risk analyses of engineered systems commonly rely on probabilistic descriptions of component reliabilities in combination with models of system performance. This approach has proven effective in determining (small) probabilities of (extreme) system failure events. However, in many industrial and academic applications of the theory, independence among component (sub-system) performances is assumed, mainly for two reasons: (a) data for determining statistical dependence is sparse, and (b) the inclusion of statistical dependence into the analysis leads to more demanding computations. In recent years, Bayesian networks (BNs) have become popular in risk analysis, partly because they provide an answer to these challenges. The graphical structure of a BN is an intuitive tool to consistently describe dependence among random variables when a purely data-driven approach is infeasible. By relying on conditional independence assumptions, it has the potential to represent dependence efficiently, i.e. with smaller data requirements and reduced computation. This not least facilitates the use of the models in the context of (near-)real-time risk management.

In this contribution, I will review recent applications of BNs for risk analysis, and on this basis discuss fundamental challenges and opportunities faced in this domain. Application examples include transportation infrastructure risk management [1], warning systems and emergency response for natural hazards [2,3], and risk assessment of technical systems with human interaction [3]. Through these examples I will demonstrate how the model building process is supported by the graphical structure and highlight the importance of fast inference in these models. I will also discuss the limitations of the approach, which are mostly associated with dependence structures that do not facilitate exploiting conditional independence.

References:

- [1] Bensi M.T., Der Kiureghian A., Straub D. (2015). Framework for Post-Earthquake Risk Assessment and Decision Making for Infrastructure Systems. *ASCE-ASME Journal of Risk and Uncertainty Analysis. Part A: Civ. Eng.* **1**(1), 04014003.
- [2] Frey D., Butenuth M., Straub D. (2012): Probabilistic graphical models for flood state detection of roads combining imagery and DEM. *IEEE Geoscience and Remote Sensing Letters* **9**(6):1051-1055.
- [3] Sättele M., Bründl M., Straub D. (2015). Reliability and Effectiveness of Warning Systems for Natural Hazards: Concepts and Application to Debris Flow Warning. *Reliability Engineering & System Safety* **142**:192-202.
- [4] Zwirgmaier K., Straub D., Groth K. (2016) Capturing cognitive causal paths in human reliability analysis with Bayesian network models. *Reliability Engineering & System Safety*, under review.

Bernd Sturmfels, UC Berkeley and MPI-MIS Leipzig, Germany

CONVEXITY IN TREE SPACES

Abstract: We study the geometry of metrics and convexity structures on the space of phylogenetic trees, here realized in terms of ultrametrics. The $\text{CAT}(0)$ -metric of Billera-Holmes-Vogtman arises from the theory of orthant spaces. While its geodesics can be computed by the Owen-Provan algorithm, geodesic triangles are complicated. We show that the dimension of such a triangle can be arbitrarily high. Tropical convexity and the tropical metric behave better. They exhibit properties desirable for geometric statistics, such as geodesics of small depth.

This talk is based on a joint work with Bo Lin, Xiaoxian Tang, and Ruriko Yoshida.

References:

- [1] S. L. Lauritzen (1996). *Graphical Models*. Oxford University Press, Oxford.
- [2] Y. Wang and S. A. Stoev (2011) Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability* **43**(2): 461–483.

Caroline Uhler, Massachusetts Institute of Technology, USA

LEARNING DAGs BASED ON SPARSEST PERMUTATIONS

Abstract: We consider the problem of learning a Bayesian network or directed acyclic graph (DAG) model from observational data. We propose the sparsest permutation algorithm, a nonparametric approach based on finding the ordering of the variables that yields the sparsest DAG. We prove consistency of this method under strictly weaker conditions than usually required. We discuss how to find the sparsest ordering by introducing the DAG associahedron and a simplex-type algorithm on this convex polytope. We end with discussing some applications to estimating gene regulatory networks.

Nanny Wermuth, Chalmers University of Technology, Gothenburg, Sweden and Johannes Gutenberg-University Mainz, Germany

ON PALINDROMIC ISING MODELS WITH GRAPH STRUCTURE

Abstract: An example of a palindromic sentence, which respects the spacings between words, is "step on no pets". It gives the same sentence when read in reverse order. This symmetry notion is applied to Bernoulli distributions. Palindromic Bernoulli distributions have uniform margins and their covariance matrix coincides with the correlation matrix. More importantly, they are now known to be characterized by missing odd-order interactions, no matter whether these are of the linear, log-linear or multivariate logistic type.

For Ising models with palindromic structure, there are at most two-factor log-linear interactions and no log-linear main effects, so that the vanishing of a single canonical parameter shows — just as in joint Gaussian distributions — conditional independence given all remaining variables. In addition, this is often, but not always, equivalent to a graph-structured partial correlation matrix, a standardized version of the inverse covariance matrix.

In this lecture, I also concentrate on additional special features which arise when the partial correlation matrices have further simplified structure. This relates to previous results by Hoesgaard and Lauritzen (2008).

This talk is based on joint work with Giovanni Marchetti, University of Florence.

References:

- [1] Fallat, S., Lauritzen S., Sadeghi K., Uhler, C., Wermuth N. and Zwiernik, P. (2016). Total positivity in Markov structures. *Ann. Statist.*, to appear; also on ArXiv: 1510.01290.
- [2] Højsgaard, S. and Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. Roy. Statist. Soc. B* **70** (5): 1005–1027.
- [3] Loh, P. and Wainwright, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Ann. Statist.*, **41**: 3022–3049.
- [4] Marchetti, G.M. and Wermuth, N. (2016). Palindromic Bernoulli distributions. *Electronic J. Statist.*, to appear; also on ArXiv: 1510.09072.

Piotr Zwiernik, Pompeu Fabra University, Barcelona, Spain

THE CORRELATION SPACE OF GAUSSIAN LATENT TREE MODELS AND MODEL SELECTION WITHOUT FITTING

Abstract: In phylogenetics and linguistics latent tree models are used to model evolutionary processes. Model selection procedures are employed to choose the best tree fitting the data. However, deciding if the tree hypothesis is consistent with the data is typically hard. We provide the full semialgebraic description of Gaussian latent tree models and link them to phylogenetic oranges. We then use this geometric description to propose a quick and robust way of choosing the best tree, or, of testing the tree hypothesis.

This talk is based on a joint work with John Aston, Nat Shiers, and Jim Smith.

References:

- [1] N. Shiers, P. Zwiernik, J. Aston, J. Smith. The correlation space of Gaussian latent tree models and model selection without fitting, *to appear in Biometrika*.
- [2] M. Drton, H. Massam, I. Olkin. Moments of minors of Wishart matrices. *Ann. Statist.* **36**, 2008.

Abstracts of Posters

Carlos Améndola, Technical University of Berlin, Germany

MAXIMUM NUMBER OF MODES OF GAUSSIAN MIXTURES

Abstract: Gaussian mixture models are widely used in Statistics, and in particular can be taken naturally as prior and posterior distributions for conditionally independent observations. A fundamental aspect of a Gaussian mixture density is its number of local maxima, or modes. We expand on known results about the number of modes that can arise when mixing k Gaussians in \mathbb{R}^d and give new bounds on their maximum number.

This poster is based on joint work with Christian Haase and Alexander Engström.

References:

- [1] C. Améndola, A. Engström and C. Haase (2016) Maximum number of modes of Gaussian mixtures. *[In preparation]*
 - [2] S. Ray and B. Lindsay (2005) The topography of multivariate normal mixtures. *Annals of Statistics* **33**: 2042-2065.
 - [3] S. Ray and D. Ren (2012) On the upper bound of the number of modes of a multivariate normal mixture. *Journal of Multivariate Analysis* **108**: 41-52.
-

Carsten Chong, Technical University of Munich, Germany

CONTAGION IN FINANCIAL SYSTEMS: A BAYESIAN NETWORK APPROACH

Abstract: We develop a structural default model for interconnected financial institutions in a probabilistic framework. For all possible network structures we characterize the joint default distribution of the system using Bayesian network methodologies. Particular emphasis is given to the treatment and consequences of cyclic financial linkages. We further demonstrate how Bayesian network theory can be applied to detect contagion channels within the financial network, to measure the systemic importance of selected entities on others, and to compute conditional or unconditional probabilities of default for single or multiple institutions.

This poster is based on a joint work with Claudia Klüppelberg.

References:

- [1] Chong, C., and Klüppelberg, C.: Contagion in Financial Systems: A Bayesian Network Approach. Submitted, 2016, Preprint available under SSRN ID 2804291.

Matthias Eckardt, Humboldt University of Berlin, Germany

GRAPHICAL MODELLING OF SPATIAL POINT PATTERNS

Abstract: In this talk we propose a novel graphical model, termed the spatial dependence graph model, which captures the global dependence structure of different events that occur randomly in space. In the spatial dependence graph model, the edge set is identified by using the conditional partial spectral coherence. Thereby, nodes are related to the components of a multivariate spatial point process and edges express orthogonality relation between the single components. We introduce an efficient approach towards pattern analysis of highly structured and high dimensional spatial point processes. Unlike all previous methods, our new model permits the simultaneous analysis of all multivariate conditional interrelations. The potential of our new technique to investigate multivariate structural relations is illustrated using data on forest stands in Lansing Woods as well as monthly data on crimes committed in the City of London.

References:

- [1] M. Eckardt. Graphical modelling of spatial point patterns. *Submitted*.
 - [2] M. A. Muggleston and E. Renshaw, E. (1996a). The exploratory analysis of bivariate spatial point pattern using cross-spectra. *Environmetrics* **7**: 361–377.
-

Eva Endres, Ludwig Maximilians University of Munich , Germany

STATISTICAL MATCHING OF DISCRETE DATA BY BAYESIAN NETWORKS

Abstract: Statistical matching (also known as data fusion, data integration, or data merging) aims at the combination of two or more data sets with only partially overlapping variable sets and disjoint sets of observations to achieve joint information about not jointly observed variables (e.g., [1]). This means that there are, for example, three sets of variables \mathbf{X} , \mathbf{Y} , and \mathbf{Z} , where $\{\mathbf{X}, \mathbf{Y}\}$ have been observed in data set A , and $\{\mathbf{X}, \mathbf{Z}\}$ in data set B . Many statistical matching procedures are based on the assumption that \mathbf{Y} and \mathbf{Z} are conditionally independent given \mathbf{X} . These conditional independence assumptions can be represented and extended by utilizing probabilistic graphical models (e.g., [4]). In this poster we present an approach of statistical matching of discrete data with the aid of Bayesian networks [2]. An exemplary application of this approach to the GGSS data [3] is shown and several ideas for extensions and further research in this area are presented.

This poster is based on a joint work with Thomas Augustin.

References:

- [1] M. D'Orazio, M. Di Zio, and M. Scanu (2006). *Statistical Matching: Theory and Practice*. Wiley, Chichester, United Kingdom.
- [2] E. Endres and T. Augustin (2016) Statistical Matching of Discrete Data by Bayesian Networks. To appear in *Proceedings of the eighth International Conference on Probabilistic Graphical Models*.
- [3] GESIS – Leibniz Institute for the Social Sciences (2013). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2012/German General Social Survey GGSS 2012*. ZA4614 Data file Version 1.1.1, doi:10.4232/1.12209.
- [4] D. Koller and N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA.

Péter Kevei, Technical University of Munich, Germany

ASYMPTOTIC MOVING AVERAGE REPRESENTATION OF HIGH-FREQUENCY SAMPLED
MULTIVARIATE CARMA PROCESSES

Abstract: High-frequency sampled multivariate continuous time autoregressive moving average processes are investigated. We obtain asymptotic expansion for the spectral density of the sampled MCARMA process $(Y_{n\Delta})_{n \in \mathbb{Z}}$ as $\Delta \downarrow 0$, where $(Y_t)_{t \in \mathbb{R}}$ is an MCARMA process. We show that the properly filtered process is a vector moving average process, and determine the asymptotic moving average representation of it, thus generalizing the results by Brockwell et al. [1,2] in the univariate case to the multivariate model. The determination of the moving average representation of the filtered process, important for the analysis of high-frequency data, is difficult for any fixed positive Δ . However, the results established here provide a useful and insightful approximation when Δ is very small.

References:

- [1] P. J. Brockwell, V. Ferrazzano, and C. Klüppelberg (2012) High-frequency sampling of a continuous-time ARMA process. *J. Time Series Anal.*, 33(1):152–160.
 - [2] P. J. Brockwell, V. Ferrazzano, and C. Klüppelberg (2013) High-frequency sampling and kernel estimation for continuous-time moving average processes. *J. Time Series Anal.*, 34(3):385–404.
-

Jürgen Landes, Ludwig Maximilians University of Munich , Germany

VARIETY OF EVIDENCE

Abstract: The Variety of Evidence Thesis is widely considered to be a truism of scientific methodology. Its probabilistic character makes it highly amenable for a Bayesian analysis. Curiously, recently put-forward Bayesian models of scientific inquiry violate versions of the Variety of Evidence Thesis. We here put forward Bayesian models of scientific inquiry in which we explicate variety of evidence and subsequently show that our explication of the Variety of Evidence Thesis holds in all our models. Our models also pronounce on disconfirmation and discordant evidence. We argue that our models pronounce rightly. Out of this discussion, the case for the Variety of Evidence Thesis emerges strengthened.

Varied evidence speaking in favor of a hypothesis confirms it more strongly than less varied evidence, *ceteris paribus*. This Variety of Evidence Thesis (VET) can be found in [5, p. 118] “*It is an undeniable element of scientific methodology that theories are better confirmed by a broad variety of different sorts of evidence than by a narrow and repetitive set of data*”, in [4, p. 77] “*It is a truism of scientific methodology that variety of evidence can be as important or even more important than the sheer amount of evidence*” and in [3, p. 94] “*Seeking a variety of evidence for a hypothesis is standard practice in science, as well as in normal life*”. [3, p. 95] offers the following semi-formal definition

Variety-of-evidence thesis. *Ceteris paribus, the strength of confirmation of a hypothesis by an evidential set increases with the diversity of the evidential elements in that set.*

In the 2000’s, Bovens & Hartmann put forward Bayesian models of scientific inference, see [1, 2]. Key to their approach is that they “*interpret more varied evidence as evidence that stems from multiple instruments (rather than a single instrument) and that tests multiple testable consequences (rather than a single testable consequence) of the hypothesis*”, see [2, p. 93-94]. Their analysis unearths cases in which a version of the VET fails, within their models.

Recently, Claveau used a different way to model reliability of (scientific) instruments. Rather than treating instruments as randomisers – the approach taken by Bovens & Hartmann – he modelled instruments as possibly systematically biased. His models satisfy a version of the VET on first pass, see [3, Section 4].¹ On second pass, Claveau drops the assumption that the systematic biases of instruments are fully independent. In this second class of models in which systematic biases of instruments are dependent to some degree, his version of the VET fails in certain cases, see [3, Section 5]. Seemingly puzzled by his findings he ends his discussion with ([3, p. 113]):

¹This is consistent with the Bovens & Hartmann approach since their model “does not apply to unreliable instruments that do not randomize”, see [2, p. 95].

The fate of the variety-of-evidence thesis is not yet settled.

Those who hold the VET dear and are thus deeply troubled by the negative results obtained by Bovens & Hartmann and Claveau face the challenge of squaring their bond to the VET and these negative results which seemingly spell an end of the VET. We will meet this challenge here by devising models of scientific inference in which a version of the VET holds. The plan for battle is as follows: First, we need to model scientific inference in which it makes sense to think about variety of evidence. We hence put forward such Bayesian net models. We can then explicate a notion of *graded variety* of a body of evidence. This puts us in a position to offer an explication of the VET within these models. The final step in this argument is to show that the VET holds in all our models.

Our models also pronounce on – rightly we think – relative disconfirmation and discordant evidence.

Out of this discussion, the case for the Variety of Evidence Thesis emerges strengthened.

References:

- [1] Luc Bovens and Stephan Hartmann. Bayesian Networks and the Problem of Unreliable Instruments. *Philosophy of Science*, 69(1):29-2, 2002.
- [2] Luc Bovens and Stephan Hartmann. *Bayesian Epistemology*. Oxford University Press, 2003.
- [3] François Claveau. The Independence Condition in the Variety-of-Evidence Thesis. *Philosophy of Science*, 80(1):94-118, 2013.
- [4] John Earman. *Bayes or Bust?* MIT Press, 1992.
- [5] Paul Horwich. *Probability and Evidence*. Cambridge University Press, 1982.

Dominik Müller, Technical University of Munich, Germany

REPRESENTING SPARSE GAUSSIAN DAGs AS SPARSE R-VINES ALLOWING FOR NON-GAUSSIAN DEPENDENCE

Abstract: Modeling dependence in high dimensional systems with parsimonious models has become increasingly important for natural and social sciences. Most approaches rely on assuming a multivariate normal distribution such as statistical models on directed acyclic graphs (DAGs). These are based on modeling conditional independencies and are scalable to high dimensions, where sparsity of the model corresponds to sparsity of the respective graph. In contrast, vine copula based dependence models can accommodate more elaborate features like asymmetry and tail dependence, where the marginal distributions can be chosen completely arbitrarily. This flexibility comes however at the cost of exponentially increasing complexity for model selection and estimation. To ease this, we introduce a novel connection between sparse DAG models and parsimonious vine copula models. Therefore, we can exploit the fast model selection and estimation of sparse DAGs while benefiting from the advantages of vine copula models. We evaluate our methodology by a large scale simulation study and high dimensional data examples demonstrating that our approach is significantly more flexible than multivariate normal models and outperforms standard methods for vine copula structure selection.

References:

- [1] Müller, D. and Czado, C. (2016), "Representing sparse Gaussian DAGs as sparse R-vines allowing for non-Gaussian dependence." *submitted for publication*, available at <http://arxiv.org/abs/1604.04202>

Moritz Otto, Karlsruhe Institute of Technology, Germany

TAIL DEPENDENCE OF REGULARLY VARYING MODELS ON DIRECTED ACYCLIC GRAPHS

Abstract: We investigate multivariate regularly varying random vectors on graphs with discrete spectral measure introduced by a directed acyclic graph (DAG), which is given by a max-linear coefficient matrix. The tail dependence coefficient measures extreme dependence between two vector components, and we investigate how far it can be used to identify the full dependence structure of the random vector on a DAG or the DAG itself. Furthermore, we estimate the model based on the matrix of all tail dependence coefficients. Here we consider two scenarios. In the first case we assume that we know the DAG and estimate the max-linear coefficient matrix as well as the matrix of the edge weights of the DAG. In the second case we assume that we do not know the DAG, but observe only the multivariate data. From these observations we want to infer the causal dependence structure in the data.

This poster is based on joint work with Nadine Gissibl and Claudia Klüppelberg.

References:

- [1] Gissibl, N. and Klüppelberg, C. (2015). Max-linear models on directed acyclic graphs. URL <http://arxiv.org/abs/1512.07522>.
- [2] Gissibl, N., Klüppelberg, C. and Otto, M. (2016). Tail dependence of regularly varying models on directed acyclic graphs. *In preparation*.

Elisa Perrone, IST Austria, Klosterneuburg, Austria

GEOMETRY OF DISCRETE COPULAS FOR WEATHER FORECASTING

Abstract: In this work we present a geometric approach to describe families of discrete copulas through the properties of their associated polytopes. In particular, we here study the polytope of ultramodular discrete copulas, i.e., discrete restrictions of ultramodular copulas. We introduce the notion of ultramodularity in the discrete setting and characterize the polytope of ultramodular discrete copulas by its defining half-spaces. We further discuss the volume of this polytope and its vertices. Finally, we present an application of the proposed approach to weather forecasting problems. Specifically, we show how the limitations of empirical postprocessing methods of numerical weather ensemble forecasts in the presence of ties can be overcome by understanding the geometry of the space of discrete copulas.

This poster presentation is based on a joint work with Caroline Uhler (MIT, Cambridge, USA).

References:

- [1] E. P. Klement, M. Manzi, and R. Mesiar (2014). Ultramodularity and copulas. *Rocky Mountain Journal of Mathematics*, **44**(1):189–202.
- [2] A. Kolesarova, R. Mesiar, J. Mordelova, and C. Sempi (2006). Discrete Copulas. *IEEE Transactions on Fuzzy Systems*, **14**(5):698–705.
- [3] R. B. Nelsen (2006). *An Introduction to Copulas*. Springer, 2nd edition.
- [4] R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, **28**(4):616–640.
- [5] G. M. Ziegler (1995). *Lectures on polytopes*. Springer-Verlag, New York.

Viet Son Pham, Technical University of Munich, Germany

VOLTERRA-TYPE ORNSTEIN-UHLENBECK PROCESSES IN SPACE AND TIME

Abstract: We propose a novel class of tempo-spatial Ornstein-Uhlenbeck processes as solutions to Lévy-driven Volterra equations with additive noise and multiplicative drift. After formulating conditions for the existence and uniqueness of solutions, we derive an explicit solution formula and discuss distributional properties such as stationarity, second-order structure and short versus long memory. Furthermore, we analyze in detail the path properties of the solution process. In particular, we introduce different notions of càdlàg paths in space and time and establish conditions for the existence of versions with these regularity properties. The theoretical results are accompanied by illustrative examples.

This poster is based on a joint work with Carsten Chong.

Roland Poellinger, Ludwig Maximilians University of Munich, Germany

VARIABLE ENTANGLEMENT IN BAYES NET CAUSAL MODELS

Abstract: In many cases of causal reasoning non-causal, non-directional knowledge is drawn on and computed efficiently and consistently (see [7]), although reasoning with this kind of knowledge seems to violate the causal Markov condition in standard Bayes net causal models (as elaborately presented, e. g., in [4, 5, 6, 8]). Embedding "entangled" variables in causal models renders those models non-Markovian. Much-discussed examples are found in causal decision theory, where modeling Newcomb-style paradoxes in standard Bayes net causal models seemingly yields counter-intuitive solutions (see [2, 3]). In this talk, I will focus on one distinguished type of "variable entanglement", namely deterministic, non-causal, and non-directional relations (called *epistemic contours* or *ECs*), formalized as 1-1 functions in extensions of deterministic structural Bayes net causal models, general causal knowledge patterns (*CKPs*).

Propagating information instantaneously, ECs exactly mark those variables in a model that cannot be modified separately. In particular, interventions will not determine causal directionality. Introducing such contours consequently violates the Markov assumption by invalidating the screening-off property of variables in a Bayesian network. For consistent reasoning to remain possible, at all, the concept of independence, as expressed in the graphical d-separation criterion, has to be extended. An additional graphical criterion, the principle of explanatory dominance, is needed to define under which conditions Markov can be reclaimed and *CKPs* utilized for causal inference. Structure alone will not suffice for this task - more information is needed and comes in the form of intensional defaults and deviants as discussed, e. g., by *Hitchcock* ([1]). Finally, to determine in which contexts which sub-portions of a partially directed graph support causal inference, the concept of identifiability of causal effects (on epistemic contours) will be modified suitably.

References:

- [1] Hitchcock, Christopher. 2007. Prevention, Preemption, and the Principle of Sufficient Reason *Philosophical Review*, 116, 495–532.
- [2] Lewis, David. 1979. Prisoners' Dilemma is a Newcomb Problem. *Philosophy & Public Affairs*, 8(3), 235–240.
- [3] Nozick, Robert. 1969. Newcomb's Problem and Two principles of Choice. Pages 114–146 in: Rescher, Nicholas (ed), *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel.
- [4] Pearl, Judea. 1995. Causal Diagrams for Empirical Research. *Biometrika*, 82(4), 669–688.

- [5] Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
 - [6] Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning. MIT Press.
 - [7] Williamson, Jon. 2009. Probabilistic Theories. In: *The Oxford Handbook of Causation*, Chap. 9, pages 185–212. Oxford University Press.
 - [8] Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation* (Oxford Studies in the Philosophy of Science). Oxford University Press.
-

Thiago do Rego Sousa, Technical University of Munich, Germany

INDIRECT INFERENCE FOR LÉVY-DRIVEN ORNSTEIN UHLENBECK (OU) PROCESSES

Abstract: Lévy-driven Ornstein-Uhlenbeck (OU) processes have been analysed in Barndorff-Nielsen and Shephard (2001) concerning both theoretical aspects and their application in the context of stochastic volatility modelling in finance. In certain situations standard likelihood-based estimation can not be applied, and the Indirect Inference Method (IIM) developed in Gouriéroux et al. (1993) has proved to be a simple and useful alternative to estimate the parameters of models defined through stochastic differential equations (SDE). IIM only requires that the model can be simulated and a reasonable estimation method for an approximate model.

In this context, Raknerud and Skare (2009) applied IIM to non-Gaussian-driven stochastic volatility models based on OU processes with a Gaussian state-space model as approximation. Bayraci and Ünal (2014) estimated the parameters of the Lévy-driven continuous time GARCH (COGARCH(1,1)) model by IIM with a discrete time GARCH(1,1) model as approximation.

In this work we study IIM for a stochastic volatility model based on an OU process driven by a compound Poisson process with GARCH(1,1) process as auxiliary model. Our Monte Carlo experiments show that the IIM estimator with auxiliary GARCH(1,1) performs similar to the estimation based on the Gaussian state-space model.

This poster is based on a joint work with Claudia Klüppelberg

References:

- [1] Barndorff-Nielsen, Ole E. and Shephard, Neil (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63.2**: 167-241.
- [2] Bayraci, Selcuk and Gazanfer, Ünal (2014) Stochastic interest rate volatility modeling with a continuous-time GARCH (1, 1) model. *Journal of Computational and Applied Mathematics* **259**: 464-473.
- [3] Gouriéroux, Christian, Monfort, Alain and Renault, Eric (1993) Indirect inference. *Journal of applied econometrics* **8.S1**: S85-S118.
- [4] Raknerud, Arvid and Skare, Øivind (2009) Indirect inference methods for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes. *Discussion Papers* **601**, Statistics Norway.

Alexander Wolf, Institute of Computational Biology, Helmholtz Munich, Germany

CAUSAL INFERENCE OF GENE REGULATION FROM SINGLE-CELL EXPRESSION DATA

Abstract: Can we qualitatively improve the inference of gene regulatory networks when using single-cell instead of bulk snapshots of gene expression?

We use a recently proposed diffusion-map based algorithm for approximately reconstructing a temporal ordering on a data manifold of single-cell measurements [1]. This amounts to inferring - from static snapshot data alone - how cells progress through a dynamic biological process. By that, initially hidden dynamic information associated with the data manifold can be revealed.

Using the reconstructed temporal ordering on the manifold, we find causal interactions among genes by extending an approach termed convergent cross mapping [2]. We explain why this is superior to other analyses strategies for observational data, which are based on cross correlation, Granger causality, Transfer Entropy, the PC algorithm and Structural equation models. Our method obtains a directed network and is efficient and scalable. We illustrate the power of the method for artificial and real data for blood development in mouse [3] and show that it substantially exceeds the predictive power of state of the art inference algorithms for gene regulatory networks.

References:

- [1] Haghverdi, L., Buettner, M., Wolf, F.A., Buettner, F. and Theis, F.J. (2016). Diffusion pseudotime robustly reconstructs lineage branching.
URL <http://biorxiv.org/content/early/2016/03/02/041384>. *Accepted in Nature Methods*.
- [2] Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M. and Munch, S. (2012). Detecting causality in complex ecosystems. *Science* **383**(6106):496–500.
- [3] Moignard, V. et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnology* **33**:296–276.