Workshop on

"High Dimensionality and Data Analysis"

September 17, 2021

organized by
Lena Janys (University Bonn), Dominik Liebl (University Bonn)

Abstracts

Mathias Drton (Technical University of Munich)

Causal Discovery with Graphical Models

Abstract: Causal discovery is the problem of inferring cause-effect relationships among a set of variables on the basis of multivariate data, where these variables are jointly observed. Common methods to tackle this problem are based on directed graphical models, which are able to tractably capture stochastic dependencies resulting from the causal relations. In this framework, methods for causal discovery solve the model selection problem of inferring the graph underlying the directed graphical model. In the talk, I will review key ideas in causal discovery, such as exploiting conditional independences among the variables or special functional forms of causal relations. The ideas will be exemplified via recent projects that develop methods to cope with the presence of latent variables and methods to supply statistical confidence statements for discovered causal effects. The specific focus will be on a local version of popular algorithms that test conditional independences, and methods based on linear non-Gaussian models.

Siegfried Hörmann (Technical University of Graz)

Preprocessing Functional Data by a Factor Model Approach

Abstract: We consider functional data which are measured on a discrete set of observation points. Often such data are measured with noise, and then the target is to recover the underlying signal. Commonly this is done with some smoothing approach, e.g. kernel smoothing or spline fitting. While such methods act function by function, we argue that it is more accurate to take into account the entire sample for the data preprocessing. To this end we propose to fit factor models to the raw data. We show that the common component of the factor model corresponds to the signal which we are interested in, whereas the idiosyncratic component is the noise. Under mild technical assumptions we demonstrate that our estimation scheme is uniformly consistent. From a theoretical standpoint our approach is elegant, because it is not based on smoothness assumptions and generally permits a realistic framework. The practical implementation is easy because we can resort to existing tools for factor models. Our empirical investigations provide convincing results.

Joint work with Fatima Jammoul (TU Graz).

Angelina Roche (Université Paris Dauphine)

Sparsity in the Functional Multivariate Linear Model

Abstract: The objective of this talk is to study a sparse estimation procedure in a linear regression problem taking as input a vector of covariates which can be of different natures (vectors and/or functional data). Two Lasso type estimators will be considered and compared from a theoretical and practical point of view. This presentation will also raise the question of whether or not it is necessary to project/regularize the functional data. An application to the prediction of electricity consumption will be presented.

Johannes Lederer (Ruhr-University Bochum)

Sparse Deep Learning

Abstract: Sparsity is popular in statistics and machine learning, because it can avoid overfitting, speed up computations, and facilitate interpretations. In deep learning, however, the full potential of sparsity still needs to be explored. In this presentation, we first discuss sparsity in the framework of high-dimensional statistics and corresponding statistical theories. We then use these insights to further our understanding of sparsity in deep learning.

Simon Scheidegger (Université de Lausanne)

Global Uncertainty Quantification in a Stochastic Climate-Economy Model

Abstract: There is a growing demand to quantify parametric uncertainty as well as economic and climate uncertainty on the climate policies to tackle global warming. To investigate parametric uncertainty and nonlinear interactions among the uncertain model parameters, we, therefore, develop a high-dimensional stochastic climate-economy model that propagates parametric uncertainty as pseudo-states. We approximate all equilibrium functions using the deep learning-based global solution method. To limit the number of model evaluations to obtain convergent statistics, we further interpolate the outcomes of the cheap-to-evaluate surrogate model employing a polynomial chaos expansion, from which we analytically estimate the Sobol' indices and univariate effects. The uncertainty quantification results show that the equilibrium climate sensitivity dominates the level of the social cost of carbon. In contrast, the stochastic and persistent long-run growth risk characterizes the volatilities of economic moments.

Joint work with Felix Kübler (University of Zurich), Aleksandra Malova (Université de Lausanne), and Takafumi Usui (University of Zurich)

Gauthier Dierickx (Ruhr-University Bochum)

Testing for relevant changes in separability of space-time covariance operators

Abstract: Space-time processes are nowadays increasingly popping up in different scientific disciplines ranging from medicine over climatology to geo-statistics. To perform statistical inference on functional data sets arising in these areas, a tremendous computational power is required. This is especially true regarding the estimation of the covariance operator for space-time data living on an L^2 space. In order to be able to treat such data in an efficient way, the notion of separability, i.e., the factorization of the covariance kernel as a product of a purely space and purely time component, has been proposed. Although separable covariance operators of space-time data are widely used, this

assumption is in practice usually not satisfied. In the last decade exact hypotheses tests for separability have been derived, confirming the non-separable nature of the data. However, the user still might prefer a separable approximation. Indeed, the gains in computational efficiency might outweigh the cost of using more complicated non-separable models.
In our talk we propose a method to address this problem by considering relevant hypotheses tests. Instead of testing for exact separability, we allow small, a priori, defined deviations from separability. To achieve this purpose, we consider measures of separability built on the different separable estimators of the covariance operator and prove invariance principles for those measures. By relying on a clever self-normalization trick, our method, moreover, has the benefit of avoiding the cumbersome estimation of the long-run covariance of the limit distribution. As a further novelty, we also consider the optimal separable approximation w.r.t. the L^2-distance, generalizing Genton (2007) to the infinite dimensional case.

Joint work with Holger Dette and Tim Kutta (both Ruhr-University Bochum).

References.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18**, 681–695.

Ulrike Schneider (Technical University Vienna)

The Geometry of Model Selection and Uniqueness of Lasso-Type Methods

Abstract: We consider estimation methods in the context of high-dimensional regression models such as the Lasso and SLOPE, defined as solutions to a penalized optimization problem. The geometric object relevant for our investigation is the polytope that is dual to to the unit ball of the penalizing norm. We show that which models are accessible by such a procedure depends on what faces of the polytope are intersected by the row span of the regressor matrix. Moreover, these geometric considerations allow to derive a criterion for the uniqueness of the estimator that is both necessary and sufficient. We illustrate this approach for Lasso and SLOPE with the unit cube and the sign permutahedron as relevant polytopes.

Joint work with Patrick Tardivel (Université Bourgogne).

Christian L. Müller (LMU/Helmholtz Zentrum Munich & Flatiron Institute New York)

An Optimization Perspective on High-Dimensional Regression Problems - With Some Gut-Wrenching Applications

Abstract: In this talk, I will provide an optimization perspective on solving high-dimensional regression problems. Starting with the idea of concomitant estimation from robust statistics, I will present a unifying optimization model for maximum likelihood-type estimation (M-estimation) relying on the idea of perspective functions. The model encompasses a wide range of statistical estimators, including Huber's M-estimator, the scaled Lasso, the TREX, and constrained sparse regression and classification. I will show several applications of these models in the context of compositional data analysis of gut microbiome data. I will conclude with some recent (surprising) prospects of these models in deep learning.

Joint work with Patrick Combettes
