Joint PhD Seminar in Statistics and Stochastics

University of Bremen,
University of Hamburg,
Carl von Ossietzky University, Oldenburg

February, 28 and March, 1 2017
Bremen, Germany

Welcome

Dear attendees,

welcome to University of Bremen for the second joint PhD Seminar in Statistics and Stochastics of the universities of Bremen, Hamburg, and Oldenburg. The scope of this seminar is to bring together PhD students from the Northern German universities in the mentioned fields of study and to foster discussions among you both on topics of your research and on general PhD related questions. Moreover, the opportunity to present your results in an extended format of 35 minutes allows the colleagues from other places to get a deeper insight into current research of the other groups, and will hopefully give you valuable feedback for your work on top of your advisors'.

Speakers

Daniel Ebel, Hamburg University
Maria Mohr, Hamburg University
André Neumann, Bremen University
Arsénio Nhacolo, Bremen University
Kornelius Rohmeyer, Oldenburg/Bremen University
Natalia Sirotko-Sibirskaya, Bremen University
Eleni Vradi, Bremen University
Tino Werner, Oldenburg University

Organizing Team

Werner Brannath
Thorsten Dickhaus
Angelika May
Peter Ruckdeschel

How to reach the University of Bremen

By Car:

When approaching Bremen by road on the A1, change to the A27 in the direction of Bremen-Bremerhaven when you come to the Bremer Kreuz. Then exit the A27 at junction Universität/Horn-Lehe and follow the signs for Centrum/Universität.
The Google Maps route planner will provide you with precise directions.
There are plenty of parking lots around the Technology Park that surrounds the University and on the campus itself, although parking is subject to a charge (70 Euro cents per day) that has to be paid on entry. It is only possible to pay in cash or by debit card (have the correct money ready as not all ticketing machines are equipped to give change).

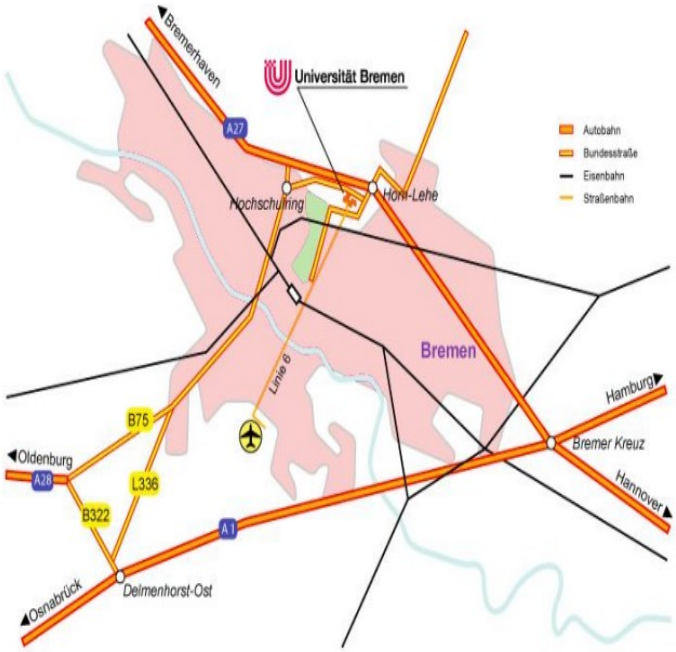
GPS: Bibliothekstraße 1, 28359 Bremen

By rail and bus:

Out and about with the Bremer Straßenbahn AG (BSAG): the following tram and bus lines run to the University main entrance: Uni Zentralbereich: 6, 20, 21, 22, 28.

Both:

At the tram station Uni Zentralbereich you cross the street for the Mehrzweckhochhaus (MZH), and there the room will be 6210, i.e., on the sixth floor.



Programme Overview

Tuesday, February 28th, 2017 (Room MZH 6210)

13:45 - 14:00 **Welcome Session**
— *Thorsten Dickhaus*, Bremen University

Block 1: Theory of multiple tests
— Chair: *Peter Ruckdeschel*, Oldenburg University

14:00 - 14:45 *Kornelius Rohmeyer*, Bremen/Oldenburg University: The Populationwise Error Rate

14:45 - 15:30 *André Neumann*, Bremen University: Estimating the Proportion of True Null Hypotheses under Dependency

15:30 - 16:00 Coffee Break

Block 2: Time series analysis
— Chair: *Angelika May*, Oldenburg University

16:00 - 16:45 *Maria Mohr*, Hamburg University: Change-point detection in a nonparametric time series regression model

16:45 - 17:30 *Natalia Sirotko-Sibirskaya*, Bremen University: A Frequency-Domain Model Selection Criterion for a (Dynamic) Factor Model

18:30 - 21:30 Conference Dinner

Wednesday, March 1st, 2017 (Room MZH 6210)

Block 3: Biometrics
— Chair: *Natalie Neumeyer*, Hamburg University

09:00 - 09:45 *Eleni Vradi*, Bremen University: Model selection based on combined penalties for biomarker identification

09:45 - 10:30 *Arsénio Nhacolo*, Bremen University: Estimation in Optimal Adaptive Phase II Oncology Trials

10:30 - 11:00 Coffee break

Block 4: Risk measures / risk minimization
— Chair: *Werner Brannath*, Bremen University

11:00 - 11:45 *Daniel Ebel*, Hamburg University: Tail-abhängige Risikomaße

11:45 - 12:30 *Tino Werner*, Oldenburg University: The ranking problem in statistical learning data

13:00 - 14:00 Farewell lunch

Abstracts

The Populationwise Error Rate - A More Liberal Error Rate for Multiplicity Adjustment in Enrichment Designs

Kornelius Rohmeyer

Bremen/Oldenburg University

14:00 - 14:45, Tuesday, February 28th, Block 1

In clinical studies control of the familywise error rate is appropriate when several hypotheses are investigated on the same population. When the population however splits into disjunct subpopulations and each hypothesis only concerns one of these without a claim beyond the subpopulation, the overall study essentially consists of separate trials which share only the same infrastructure. In this case the familywise error rate is unreasonably conservative. In some cases the subpopulations are disjunct by definition (like two groups 'biomarker positive' and 'negative/unknown'), but in many other cases the subpopulations can overlap. For this setting we propose a generalized error rate that takes into account the probability to belong to a certain subpopulation or intersection of subpopulations. This error rate - which we call the populationwise error rate - extends continuously the spectrum from the FWER in the first setting to the unadjusted case for disjunct populations. We start defining simultaneous test procedures with control of the populationwise error rate. We then generalize the closed testing principle and show how to construct step-down tests. The gain in power and sample size by using the populationwise error instead of the familywise error rate is illustrated by first examples.

Estimating the Proportion of True Null Hypotheses under Dependency

André Neumann

Bremen University

14:45 - 15:30, Tuesday, February 28th, Block 1

It is a well known result in multiple hypothesis testing that the proportion π_0 of true null hypotheses is not identified under general dependencies. However, it is possible to estimate π_0 if structural information about the dependency structure among the test statistics or p-values, respectively, is available. In this talk I demonstrate these points, and explain our proposed marginal parametric bootstrap method. A pseudo-sample of bootstrap p-values is generated, which still carry information about π_0, but behave like realizations of stochastically independent random variables. Theoretical properties of resulting estimation procedure for π_0 will be briefly discussed and their usage will be illustrated in computer simulations.

Change-point detection in a nonparametric time series regression model

Maria Mohr

Hamburg University

16:00 - 16:45, Tuesday, February 28th, Block 2

A weakly dependent time series $(X_t, Y_t)_{t\in\mathbb{Z}}$ in $\mathbb{R}^d \times \mathbb{R}$ is considered, for which we develop a strategy to detect whether the nonparametric conditional mean function $m_t(\cdot) = E[Y_t|X_t = \cdot]$ is stable in time $t \in \mathbb{Z}$. The strategy also allows for heteroscedasticity. Our proposal is based on a modified CUSUM type test procedure, which uses a sequential marked empirical process of residuals. Empirical process theory is required to show weak convergence of the considered process to a centered Gaussian process under the null „$m_t(\cdot) = m(\cdot)$ *for all* t“ and a stationarity assumption. As a consequence we obtain the convergence of the Kolmogoroff-Smirnoff and Cramér-von Mises type test statistics. The proposed procedure acquires a very simple limiting distribution and nice consistency properties against change-point alternatives, features from which related tests are lacking. A simulation study is conducted to investigate the finite sample performance of our test.

A Frequency-Domain Model Selection Criterion for a (Dynamic) Factor Model

Natalia Sirotko-Sibirskaya

Bremen University

16:45 - 17:30, Tuesday, February 28th, Block 2

We consider a multivariate time series model of the form

$$\mathbf{X}(t) = \sum_{s=-\infty}^{\infty} \Lambda(s)\,\mathbf{f}(t-s) + \boldsymbol{\varepsilon}(t),\ 1 \le t \le T,$$

denoted a dynamic factor model as a p-dimensional, covariance-stationary stochastic process in discrete time with mean zero, $\mathbf{X} = (\mathbf{X}(t) : 1 \le t \le T)$, can be decomposed into a k-dimensional vector of so-called common factors, $\mathbf{f}(t) = (f_1(t), \ldots, f_k(t))^\top$, and a p-dimensional vector of "specific" or "idiosyncratic" factors, $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_p(t))^\top$. This decomposition is used when it is assumed that $k << p$, therefore, the key step is in finding the rank of $\Lambda(s)\,\mathbf{f}(t-s)$ (for each s), so that the dynamics of the process itself, i.e. in $\mathbf{X}$, can be recovered by such lower-rank representation.
Up to now there is no unanimous agreement among the researchers on which method to use in order to choose the optimal number of factors. Classical methods include computing likelihood-ratio tests and using screeplots in principal-component analysis, however, these methods impose an assumption of homoskedastic noise of idiosyncratic factors which can be regarded as limiting in view of the current research. Recent developments include AIC/BIC type of criteria adapta-

tion to factor model analysis, see Bai and Ng (2002), dynamic principal component analysis, see Hallin and Liska (2007), bi-cross-validation, see Owen and Wang (2005), and others.
We propose a data-driven method for selecting an "optimal" number of factors. The method is based on cross-validation technique for a factor model evaluated in the frequency domain and allows to relax the assumption of homoskedasticity of idiosyncratic components. In the spirit of Hurvich and Zeger (1990) we define a frequency-domain-cross-validation criterion, in this case, for a factor model. It can be shown that expectation of a frequency-domain cross-validation criterion is approximately equal to the sum of the MSE of a spectrum estimate and variance of idosyncratic components. This criterion is evaluated for each possible choice of k. The choice of the "optimal" model is based on minimization of the corresponding criterion. The proposed method is then compared to several existing criteria in Monte-Carlo simulations as well as applied to a data set to evaluate its performance empirically.

Model selection based on combined penalties for biomarker identification (Coauthors: Werner Brannath, Thomas Jaki (Lancaster University), Richardus Vonk (Bayer AG))

Eleni Vradi

Bayer AG/Berlin and Bremen University

09:00 - 09:45, Wednesday, March 1th, Block 3

The growing role of targeted medicine has led to an increased focus on the development of actionable biomarkers. Current penalized selection methods that are used to identify biomarker panels for classification in high dimensional data, however, often result in highly complex panels that need careful pruning for practical use. In the framework of regularization methods a penalty that is a weighted sum of the L_1 and L_0 norm has been proposed to account for the complexity of the resulting model. In practice, the limitation of this penalty is that the objective function is non-convex, non-smooth, the optimization is computationally intensive and the application to high-dimensional settings is challenging. In this paper we propose a stepwise forward variable selection method which combines the L_0 with L_1 or L_2 norms. The penalized likelihood criterion that is used in the stepwise selection procedure results in more parsimonious models, keeping only the most relevant features. Simulation results and a real application show that our approach exhibits a comparable performance with common selection methods with respect to the prediction performance whilst minimizing the number of variables in the selected model resulting in a more parsimonious model as desired.

Estimation in Optimal Adaptive Phase II Oncology Trials

Arsénio Nhacolo

Bremen University

09:45 - 10:30, Wednesday, March 1th, Block 3

Phase II trials are concerned with making decision of whether a treatment is sufficiently efficient to worth further investigations in late large scale Phase III trials. In oncology Phase II trials, frequentist single-arm two-stage group-sequential designs with binary endpoints are commonly used. Based on ethical desirability to expose less patients to an inefficient treatment and to speed-up the development of an efficient treatment, these designs allow early termination of the trial for futility and efficiency (e.g., Schultz et al., 1973), or for futility only (e.g., Simon, 1989). Their sample sizes and decision rules for each stage are predefined. To allow flexibility, adaptive versions of these designs have been proposed. One of the recent proposals is the optimal adaptive design by Englert and Kieser (2013). Making use of discrete conditional error function (Englert and Kieser, 2012), it extends the fixed sample size oncology Phase II designs by allowing the sample size of stage two to depend on the number of responses observed in the first stage. It is optimal in a sense that it minimizes the average sample size under the null hypothesis.
Unlike in classical designs, in adaptive oncology Phase II designs the estimation of treatment effect (response rate) is not straightforward. We are going to propose interval and point estimation procedure for the optimal adaptive design. The procedure uses the concept of stage-wise ordering. Therefore, we will first propose and discuss different approaches for defining sample space ordering, from which we will derive the p-value and then interval and point estimation. We will also perform simulation studies to compare the sample space ordering approaches.

Tail-abhängige Risikomaße

Daniel Ebel

Hamburg University

11:00 - 11:45, Wednesday, March 1th, Block 4

Viele in der Praxis gebräuchliche Risikomaße hängen nur vom Tail der Verteilung des zu messenden Risikos ab, d.h. es existiert ein Funktional T, sodass man das Risikomaß für $\alpha_n \searrow 0$ schreiben kann als $T(F^{\leftarrow}(1-\alpha_n \cdot))$. In diesem Vortrag wird ein Verfahren vorgestellt, derartige Risikomaße zu schätzen. In einem ersten Schritt wird der Tail der Quantilfunktion durch die empirische Tail-Quantilfunktion Q_n ersetzt. Für das Risikomaß wird ein Schätzer basierend auf $T(Q_n)$ gewählt. Nun kann das Funktional getrennt von dem stochastischen Verhalten analytisch auf Glattheitseigenschaften (z.B. Hadamard-Differenzierbarkeit) untersucht werden. Weiterhin werden bekannte Resultate für die empirische Tail-Quantilfunktion (z.B. Drees (1998)) als Schätzer für den Tail der Quantilfunktion ausgenutzt, um asymptotische Normalität für $T(Q_n)$ nachzuweisen.

Dieses Verfahren soll am Beispiel verallgemeinerter bedingter Momente (Methni et al. (2014)) im Fall unabhängiger Daten durchgeführt werden. Darüber hinaus wird eine Entwicklung des Schätzfehlers angegeben und diskutiert.
Abschließend soll kurz das Expektil (Bellini et al. (2014)) als verallgemeinertes Quantil definiert und eine Motivation zum Schätzen dieser Größe aufgezeigt werden.

Abstract: The ranking problem in statistical learning

Tino Werner

Oldenburg University

11:45 - 12:30, Wednesday, March 1th, Block 4

In various application domains, it is not sufficient just to perform binary classification in the sense of classifying some instances in a given data set as ”good” and the others as ”bad”. Instead, one wants to find an ordering of the features to answer questions like ’Which one is the best?’. This challenge is referred to as the ”ranking problem”. It can be extended to a local variant where it is not the goal to rank all instances but only the ”good” ones. In this talk, we start from an empirical risk minimization approach provided by Clémençon et. al. to tackle the ranking problem in a general case, i.e., where the labels do not have to be binary. One will see that ranking can be regarded as pairwise binary classification. Furthermore, the main ideas of constructing suitable learning algorithms will be discussed.

Inhaltsverzeichnis

Participants

Bodnar, Rostyslav	Bremen University, Institute for Statistics	bodnar@uni-bremen.de
Brannath, Werner	Bremen University, Institute for Statistics	brannath@math.uni-bremen.de
Dickhaus, Thorsten	Bremen University, Institute for Statistics	dickhaus@uni-bremen.de
Drees, Holger	Hamburg University, Department of Mathematics	holger.drees@math.uni-hamburg.de
Ebel, Daniel	Hamburg University, Department of Mathematics	fmnv023@uni-hamburg.de
Mändle, Andreas	Oldenburg University, Institute for Mathematics	andreas.maendle@uni-oldenburg.de
May, Angelika	Oldenburg University, Institute for Mathematics	angelika.may@uni-oldenburg.de
Meyer, Christian	Oldenburg University, Institute for Mathematics	christian.meyer1989@t-online.de
Mohr, Maria	Hamburg University, Department of Mathematics	maria.mohr@uni-hamburg.de
Neumann, André	Bremen University, Institute for Statistics	neumann@uni-bremen.de
Neumeyer, Natalie	Hamburg University, Department of Mathematics	natalie.neumeyer@math.uni-hamburg.de
Nhacolo, Arsénio	Bremen University, Institute for Mathematics	anhacolo@uni-bremen.de
Rohmeyer, Kornelius	Bremen University and Oldenburg University	kornelius.rohmeyer@uni-oldenburg.de
Ruckdeschel, Peter	Oldenburg University, Institute for Mathematics	peter.ruckdeschel@uni-oldenburg.de
Sirotko-Sibirskaya, Natalia	Bremen University, Institute for Statistics	sirotkos@uni-bremen.de
Vradi, Eleni	Bayer AG, Berlin and Bremen University	eleni.vradi@bayer.com
Werner, Tino	Oldenburg University, Institue for Mathematics	tino.werner1@uni-oldenburg.de