


Joint PhD Seminar in Statistics and Stochastics

University of Bremen,

Institute for Statistics (IfS),
Competence Center for Clinicial Trials Bremen (KKSB)

Ulm University, Institute of Statistics

Carl von Ossietzky University of Oldenburg,

Institute for Mathematics,
Division of Epidemiology and Biometry (EuB)

September 30 and October 1, 2019
Bremen, Germany

Welcome

Dear attendees,

welcome to University of Bremen for the fourth joint PhD Seminar in Statistics and Stochastics, this time among the universities of Oldenburg, Ulm and Bremen. The scope of this seminar is to bring together PhD students from different 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 40 minutes including 5 to 10 minutes of discussion allows the colleagues from other places to get a deeper insight into current research of other groups, and will hopefully give you valuable feedback for your work on top of your advisors'. This year, the workshop will close with a scientific talk of our guest researcher Jan Beyersmann.

Speakers

Jan Beyersmann, Ulm University, Institute of Statistics
Jan Feifel, Ulm University, Institute of Statistics
Charlie Hillner, University of Bremen, IfS and KKSB
Anh-Tuan Hoang, University of Bremen, IfS
Julian Jetses, University of Oldenburg, Institute for Mathematics
Marius Pluhar, University of Oldenburg, Institute for Mathematics
Alexander Seipp, University of Oldenburg, EuB
Dominik de Sordi, University of Oldenburg, EuB
Regina Stegherr, Ulm University, Institute of Statistics
Jonathan von Schröder, University of Bremen, IfS
Max Westphal, University of Bremen, IfS and KKSB

Organizing Team

Werner Brannath, University of Bremen, IfS and KKSB
Markus C. Christiansen, University of Oldenburg, Institute for Mathematics
Thorsten Dickhaus, University of Bremen, IfS
Angelika May, University of Oldenburg, Institute for Mathematics
Fabian Otto-Sobotka, University of Oldenburg, EuB
Peter Ruckdeschel, University of Oldenburg, Institute for Mathematics

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 (about 1 Euro 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).

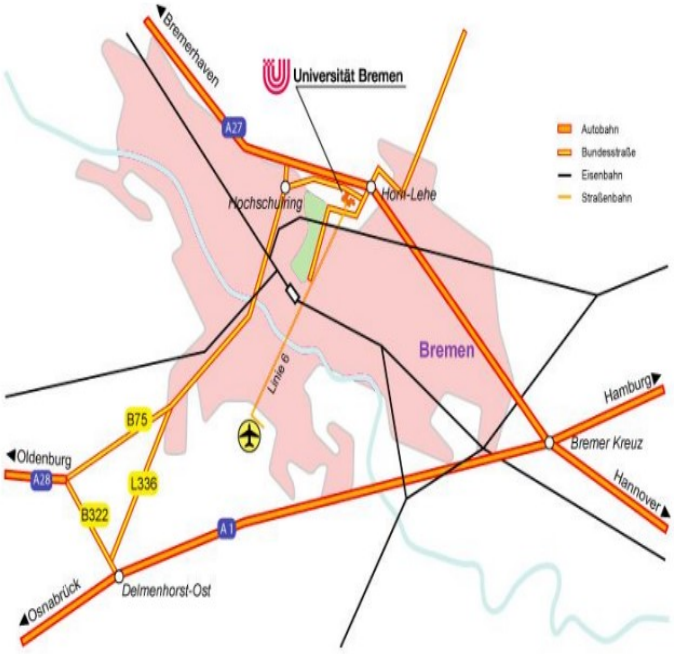
GPS: Bibliothekstraße 1, 28359 Bremen

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

Mondy, September 30, 2019 (Room MZH 6210)

09:20 - 09:30 Welcome
— *Werner Brannath*, University of Bremen

Session 1 — Chair: *Werner Brannath*, Bremen University

09:30 - 10:10 *Jonathan von Schröder*, University of Bremen (IfS):
Normalization of MALDI spectra

10:10 - 10:50 *Max Westphal*, University of Bremen (IfS & KKSB):
Bayesian (subset) selection of prediction models

10:50 - 11:10 Coffee Break

Session 2 — Chair: *Markus C. Christiansen*, Oldenburg University

11:10 - 11:50 *Jan Feifel*, Ulm University (Inst. of Statistics):
Dynamic case-control sampling designs for involved time-to event data

11:50 - 12:30 *Julian Jetses*, University of Oldenburg (Inst. of Mathematics):
Calculation of surplus participations in life and health insurance by determining martingale decompositions

12:30 - 14:10 Lunch break at University Mensa

Session 3 — Chair: *Peter Ruckdeschel*, Oldenburg University

14:10 - 14:50 *Alexander Seipp*, University of Oldenburg (EuB):
Weighted expectile regression for right-censored data

14:50 - 15:30 *Marius Pluhar*, University of Oldenburg (Inst. of Mathematics):
Time-inhomogeneous variable length Markov chains in insurance

15:30 - 15:50 Coffee break

Session 4 — Chair: *Jan Beyersmann*, Ulm University

15:50 - 16:30 *Regina Stegherr*, Ulm University (Inst. of Statistics):
Methodological aspects in safety comparisons of adverse event probabilities in time-to-event data

16:30 - 17:10 *Dominik de Sordi*, University of Oldenburg (EuB):
Handling missing continuous participant data in longitudinal studies

18:00 Dinner at Platzhirsch (Kuhgrabenweg 30, 28359 Bremen)

Tuesday, October 1st, 2019 (Room MZH 6210)

Session 5 — Chair: *Thorsten Dickhaus*, Bremen University

09:50 - 10:30 *Anh-Tuan Hoang*, University of Bremen (IfS):
On randomized p-values in replicability analysis

10:30 - 11:10 *Charlie Hillner*, University of Bremen (IfS & KKSB):
Adaptive designs with control of the population-wise error rate

11:10 - 11:30 Coffee break

Final Session — Chair: *Werner Brannath*, Bremen University

11:30 - 12:10 *Jan Beyersmann*, Ulm University (Inst. of Statistics):
Competing risks, immortal time bias: two myths in survival analysis?

12:10 - 12:20 Closing remarks *(Werner Brannath)*

12:30 Lunch at Café Unique (Enrique-Schmidt-Str. 7, 28259 Bremen)

Abstracts

Normalization of MALDI spectra

Jonathan von Schröder, jvs@uni-bremen.de
University of Bremen, Institute for Statistics

09:30 - 10:10, Monday, September 30, Session 1

Matrix Assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) is a widely used imaging mass spectrometry (IMS) technique. It is a "tool for spatially-resolved chemical analysis of diverse sample types ranging from biological and plant tissues to bio and polymer thin films" (Alexandrov et. al. 2012). When applied to mostly homogenous tissue (e.g. for classifying cancerous tissue) there is no relevant spatial structure and the individual spectra can be understood as histograms following an (unknown) distribution. However, for a MALDI observation $x \in \mathbb{R}^n$ (where n is usually of magnitude 10^3 or 10^4) the components x_i are only proportional to the number of ions observed at a certain mass to charge ratio. Thus, the exact number of ions is not known. It is however clear that the total number of ions is large (10^9) and therefore an approximation by a Gaussian distribution with an appropriate covariance structure is reasonable. Nevertheless, the number of parameters is still (at least) one magnitude larger than the number of observations making statistical modelling challenging. One basic task is the normalization of MALDI observations: While working with data sets obtained by the BMBF-MaDiPath project it became clear, that biologically very similar MALDI observations can have very different total masses when measurements are performed by different people at different locations. However, many normalization approaches lead to bin-values that are (unexpectedly) different. This talk will, therefore, describe different approaches to spectra normalization as well as numerical and statistical considerations.

Bayesian (subset) selection of prediction models

Max Westphal, mwestphal@uni-bremen.de
University of Bremen, Institute for Statistics & Competence Center for Clincial Trials Bremen

10:10 - 10:50, Monday, September 30, Session 1

A core ingredient of applied machine learning is data splitting. A strict separation between model development and evaluation is frequently advised. This recommendation can be easily implemented and allows a simple and unbiased performance assessment (estimation, hypothesis testing) for the final (prediction) model which has been trained and chosen previously on independent data. However, previous work has shown that this strategy is suboptimal, mainly because the final model choice cannot be altered without impairing the statistical inference. To resolve this issue, we proposed to evaluate several promising models simultaneously on the test data. The inference task can then be phrased as a multiple testing problem. This allows a delayed final model selection which in turn increases the final model performance and statistical power while type I errors can still (approximately) be controlled. So far, we employed different empirical subset selection rules to decide which prediction models should be evaluated. While these rules proved to be beneficial in several simulation studies, they lack a throughout theoretical justification. In this talk, we tackle the problem from the viewpoint of Bayesian decision theory. The goal

can then be phrased as to maximize the (posterior) expected utility at the time point of decision making. Of the many utility functions that have been proposed for this problem, most are however not free of 'hyperparameters' as they usually trade off the performance versus the number of the selected models. We resolve this issue by also modelling the subsequent final model selection in the evaluation study. This results in a hyperparameter free procedure which can be implemented numerically in a straightforward manner.

Dynamic case-control sampling designs for involved time-to event data

Jan Feifel, jan.feifel@uni-ulm.de
Ulm University, Institute of Statistics

11:10 - 11:50, Monday, September 30, Session 2

When investigating the effect on length-of hospital stay researchers face a sophisticated multistate model with rare intermediate and common absorbing state(s). For large cohort studies with rare outcomes nested case-control designs are favorable due to an efficient use of limited resources. If the outcome is not necessary rare, nested case-control designs are still applicable but do not truly reduce the sample size. We therefore study the nested exposure case-control design, which samples all exposed patients but not all unexposed ones. Here, the inclusion probability of observed events evolves over time. This new scheme improves on the classical nested case-control design where for every observed event controls are chosen at random. The martingale arguments underlying both designs allow for choosing controls more sophisticated than usually applied in practice. We will discuss several options how to account for past time-dependent exposure status within a nested case-control design and their related merits. It will be seen that a smart utilization of the available information at each point in time can lead to a powerful and simultaneously less expensive design. We will also sketch alternative designs, e.g. treating exposure as a left-truncation event that generates matched controls, and time-simultaneous inference of the baseline hazard using the wild bootstrap. The methods will be applied to observational data on the impact of a hospital-acquired pneumonia on the length-of-stay in hospital, which is an outcome commonly used to express both the impact and the costs of such adverse events.

Calculation of surplus participations in life and health insurance by determining martingale decompositions

Julian Jetses, julian.jetses1@uol.de
University of Oldenburg, Institute for Mathematics

11:50 - 12:30, Monday, September 30, Session 2

Insurance liabilities are influenced by various sources of risk such as equity, interest and biometric risk, so quantifying individual risk contributions is of great relevance in view of risk management. Based on the martingale representation theorem, recent literature proposed the MRT-decomposition of life insurance liabilities in a two-state model. There, the liabilities are decomposed into a sum of stochastic integrals associated with the different sources of risk. In my talk I will illustrate that this concept also gives a theoretical basis for the calculation of surplus participations. Furthermore, my talk will discuss

the MRT-decomposition for arbitrary multi-state models, where the integrands related to unsystematic risk are explicitly specified. Particularly, for modelling policyholders in an adequate multi-state framework I will introduce a pathwise definition of doubly stochastic Markov chains.

Weighted expectile regression for right-censored data

Alexander Seipp, alexander.seipp@uni-oldenburg.de
University of Oldenburg, Devision of Epidemiology and Biometry

14:10 - 14:50, Monday, September 30, Session 3

Expectile regression can be used to analyze the entire conditional distribution of a metric response, while omitting any parametric distributional assumptions. We estimate least asymmetrically weighted squares with weights defined according to the expectile level. Expectile regression generalizes conventional mean regression. Quantile regression extends median regression along the same principle. Among the benefits of expectile regression are computational simplicity, efficiency and the possibility to incorporate a semiparametric predictor. Although single expectiles are less intuitive to interpret, a set of expectiles can be converted to more easily interpreted tail expectations i.e. the expected shortfall. Because of its advantages in full data settings, we have examined extensions to right-censored data. Regression for right-censored data is often done with parametric Accelerated Failure Time Models. We see expectile regression as a distribution-free alternative that focuses on more than just mean effects. We propose to extend expectile regression with inverse probability weights. Observed cases receive weights equal to the inverse of the probability that they are uncensored, while censored observations are excluded. Estimates are easy to implement and computationally simple. Inverse probability weights have already been used to extend mean and median regression to right-censored data. We present our estimator and connections of our estimator to M-quantile regression. M-quantile regression uses asymmetric weights to generalize M-estimation in the same way as expectile regression generalizes mean regression. We show asymptotic results when the true weights are known. We discuss problems with bias at the upper tail and confidence interval coverage when the true weights are unknown. We present simulation results from an extensive simulation study in which our plug-in estimators were evaluated.

Time-inhomogeneous variable length Markov chains in insurance

Marius Pluhar, marius.pluhar1@uol.de
University of Oldenburg, Institute for Mathemtics

14:50 - 15:30, Monday, September 30, Session 3

In insurance modelling processes are often assumed to be Markovian. While the training data may not harmonize with the Markov assumption, ignoring this yields models that are easy to handle and easy to communicate to the customer. Switching over to higher order Markov models, this discrepancy can be reduced, but the model complexity of higher order models increases exponentially which makes those models difficult to handle. Within this talk the concept and an approach to fit time-inhomogeneous Variable Length Markov

Chains (tiVLMC) will be presented. tiVLMC are a subclass of Markov models that allow the model order to vary with the different conditional pasts. Differentiating between informative and non-informative pasts, tiVLMC are able to display the same dependency structure as the corresponding higher order Markov model while being less complex

Methodological aspects in safety comparisons of adverse event probabilities in time-to-event data

Regina Stegherr, regina.stegherr@uni-ulm.de
Co-Authors: Claudia Schmoor, Tim Friede, Michale Luebbert, and Jan Beyersmann
Ulm University, Institute of Statistics

15:50 - 16:30, Monday, September 30, Session 4

Safety analyses in terms of adverse events (AEs) are an important aspect of benefit-risk assessment of therapies. For time-to-event studies, AE analyses are often rather simplistic as the incidence proportion, incidence densities or a non-parametric Kaplan-Meier are used (Unkel et al. 2018). But these analyses either do not account for censoring, or rely on a too restrictive parametric model, or ignore competing risks (Allignol et al. 2016). With the non-parametric Aalen-Johansen estimator as the gold standard we investigate these potential sources of bias. As the estimators may have large variances at the end of follow-up, the estimators are not only compared at the maximal event time but also at two quantiles of the observed times. The impact of using different estimators on group comparisons is unclear, as, for example, the ratio of two both underestimating or overestimating estimators may or may not be comparable to the ratio of the goldstandard estimator. Therefore, the ratio of the AE probabilities is also calculated based on different approaches. A data example and a simulation study conduct theses comparisons under constant and non-constant hazards, different censoring mechanisms and event frequencies.

References:
Allignol A, Beyersmann J, Schmoor C (2016). Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical Statistics*, 15(4):297-305.
Unkel S, Amiri M, Benda N et al. (2018). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical Statistics*, 18:165-183.

Handling missing continuous participant data in longitudinal studies

Dominik de Sordi, dominik.de.sordi@uni-oldenburg.de
University of Oldenburg, Division of Epidemiology and Biometry

16:30 - 17:10, Monday, September 30, Session 4

In studies with repeated measurements the occurrence of missing values holds different questions. How can we find the optimal imputation method for our situation? How can we measure that optimality? Which information / covariables should be used for imputation? How do imputation methods improve the precision of the estimation? In my thesis, I deal with these questions in three steps. The first step is a systematic review, the second step are simulation studies. In the third step, the optimal method will be used in a clinical study.

In the first step of my work, I performed a systematic search to find the current methods and approaches for dealing with missing values in longitudinal data. For this review the search methods, databases and in and exclusion criteria were preset. The extraction is provided by two extractors. The relevant criteria that covered the important aspects of the methods were extracted along with the simulation settings. From the 225 papers that were accessed in full text, 115 were included in the analysis. Within these more than 100 different methods for dealing with MPD were reported. The methods used to handle MPD were categorized to multiple imputation, mixed models (40, 34.8% each), likelihood-based procedures (21, 18.3%) and pattern mixture models (19, 16.5%). The categories were not exclusive. Single imputation procedures such as last observation carried forward were often used as reference (30, 26.1%). Based on the identified methods and simulation settings I am now in the stage of planning my own simulations. The aim of my simulations is to show the advantages and disadvantages of selected methods. The scenarios will differ regarding the missing mechanisms, the total size of the data set and the amount of missing values (e.g.). It will be a challenge to create reproducible datasets, which can be used as reference for upcoming papers and by other researchers in the future. Therefore, new methods could be easily compared to my findings. As an additional point, I will include an unpublished approach that I used in my master thesis.

On randomized p-values in replicability analysis

Anh-Tuan Hoang, anhtuan.hoang@uni-bremen.de
University of Bremen, Institute for Statistics

09:50 - 10:30, Tuesday, October 1, Session 5

We consider the simultaneous testing of $m > 1$ composite null hypotheses. For these, least favorable parameter configurations (LFCs) are usually employed to compute marginal p-values, thus being over-conservative for non-LFCs. In this context, Dickhaus (2013) proposes randomized p-values, which result from a data-dependent mixing of the LFC-based p-values and a uniformly distributed random variable on [0,1] (Uni[0,1]), that is independent of the data, such that their distributions come much closer than their LFC-based counterparts to Uni[0,1]. This is especially beneficial in the estimation of the proportion of true null hypotheses. We extend the model and give a class of randomized p-values, which arise naturally in many applications. Furthermore, we give conditions for the validity of these p-values and formulas for their calculation, which turn out to be linear enlargements of their LFC-based counterparts, if the latter are small (indicating a false null hypothesis), and else uniformly distributed on [0,1]. Finally, we show how this model can be applied in replicability analysis.

Adaptive designs with control of the population-wise error rate

Charlie Hillner, chillner@uni-bremen.de
University of Bremen, Institute for Statistics and Competence Center for Clinical Trials Bremen

10:30 - 11:10, Tuesday, October 1, Session 5

In confirmatory clinical trials that concern tests of several hypotheses in several popula-

tions the multiple type-I error is usually kept small by controlling the family-wise error rate (FWER). However, if a treatment or a treatment strategy is tested in several disjoint populations, each population is effected by only a single hypothesis test. In this case, the control of the FWER might be too conservative, so a more liberal multiple type I error rate, which we denote as "population-wise error rate (PWER)", is considered. Suppose there are m possibly overlapping populations P_i in each of which the efficacy of a treatment strategy T_i is to be investigated by testing a hypothesis H_i. The population-wise error rate then describes the probability that a randomly selected future patient is assigned to an inefficient treatment strategy. In today's practice, it is common to use so-called adaptive clinical trial designs which enable us to change certain features of the design without undermining the validity of the trial, i.e. without inflating the overall type-I error rate. One way to conduct such adaptations is the Conditional Rejection Probability principle (CRP-principle) by Müller and Schäfer that makes use of the probability of erroneously rejecting the null hypothesis conditioned on the data collected so far. We propose an adaptive design strategy that ensures control of the PWER based on an adoption of the CRP-principle to the PWER-approach.

Competing risks, immortal time bias: two myths in survival analysis?

Jan Beyersmann, jan.beyersmann@uni-ulm.de
Ulm University, Institute of Statistics

11:30 - 12:10, Tuesday, October 1, Final Session

Because one has to wait for times to event to occur, survival data are incompletely observed and survival analysis is based on hazards. However, there arguably is an overemphasis on estimating survival probabilities, although hazard-based analyses are available for much more complicated situations. Two relatively straightforward extensions of a standard survival setup are time-to-first-event and type-of-first-event (aka competing risks) and time-dependent exposures and time-to-event. The overemphasis on survival curves has led to estimating cumulative event probabilities of a competing risk in a Kaplan-Meier-fashion, which inevitably overestimates. Time-dependent or immortal time bias occurs if the time-dependent exposure is analyzed as a baseline variable. Immortal time bias allows for survival curve estimation, which is otherwise complicated by the time-dependency of the exposure, but inevitably underestimates the effect of exposure in terms of the hazard ratio. We will argue that both biases could have well been banned from biostatistics ever since the heydays of William Farr (1807-1883) and Groucho Marx (1890-1977). Following Farr's example, this talk will not be very mathematical, emphasizing concepts and relying on rather simple calculations. We will demonstrate why well over 50% of all published Kaplan-Meier curves might not have a proper interpretation. Examples in the talk will include unjustified hope for metformin to improve survival also in cancer patients, why sunbathing may or may not be good for you and a recently published trial in the New England Journal of Medicine on infection prophylaxis that found that 20% is less than 5%. Research in diabetes and safety analyses will also be featured.