UNIVERSITÄT SALZBURG

10th International Workshop on Simulation and Statistics

Workshop booklet

September 2 - 6, 2019
Salzburg, Austria

Foto: Tourismus Salzburg

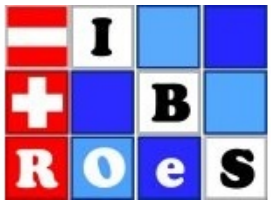
Enjoy your days in Salzburg!

Scientific Program Commitee:
Jürgen Pilz
Viatcheslav Melas
Dieter Rasch

Local organization:	**Keynote Speaker:**
Arne C. Bathke	Regina Liu
Wolfgang Trutschnig	Edgar Brunner
Georg Zimmermann	Gerd Antes
Andrea Baumgartner	Christian Robert
Martin Happ	Holger Dette


UNIVERSITÄT
SALZBURG

Fachbereich
Mathematik

ASA

I
B
R
O
e
S

OSG

BRAIN PRODUCTS
Solutions for neurophysiological research

Café
HABAKUK

Springer

LAND
SALZBURG

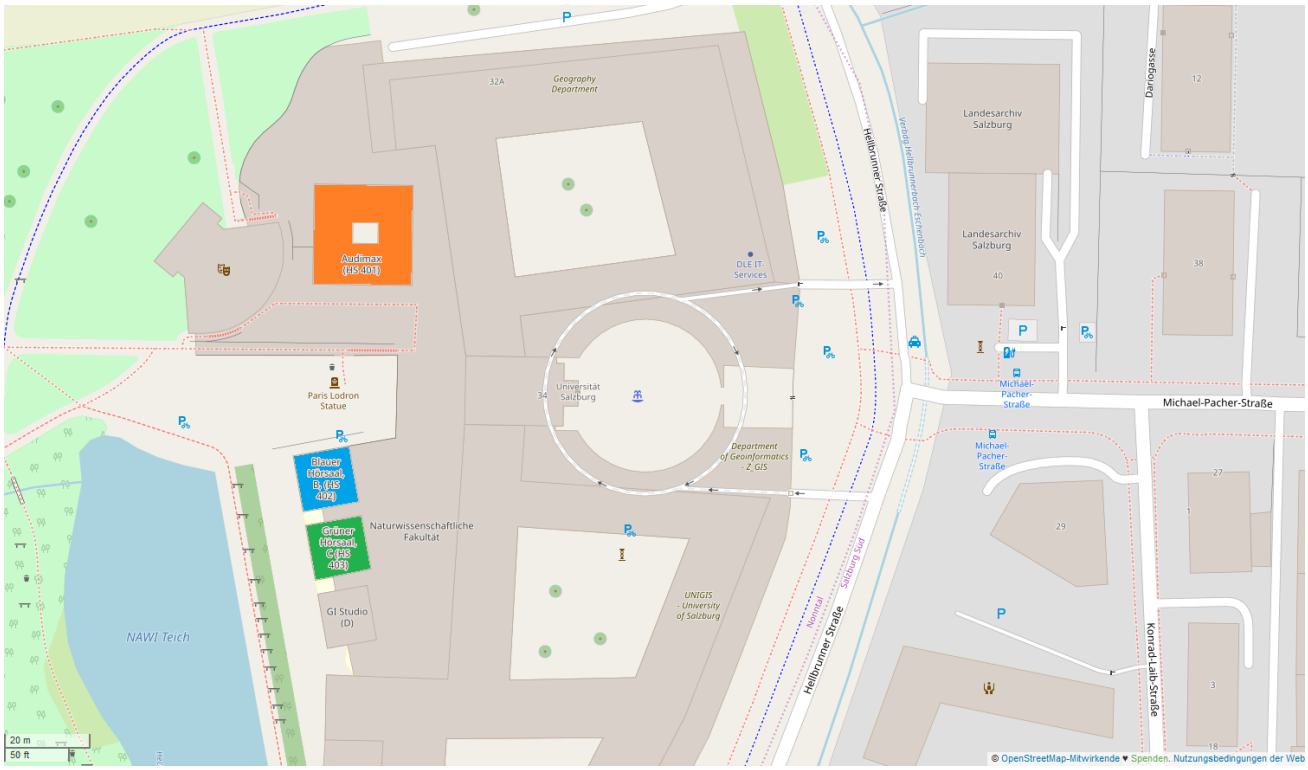
wissen:stadt
salzburg

Contents

Conference Location

The conference is taking place in the Faculty of Natural Sciences (NAWI) of the University of Salzburg in the lecture rooms 401 (Auditorium Maximum), 402 (Blauer HS) and 403 (Grüner HS). The registration desk is in front of 402 (Blauer HS).

Hellbrunner Str. 34
5020 Salzburg, Austria
Phone (Department of Mathematics): +43 662 8044 5302



https://www.openstreetmap.org/

General Information

Registration

Registration opens on Monday September 2, from 10:00 to 17:00 in front of lecture room 402 (Blue lecture room). On Tuesday to Friday from 09:00 to 17:00.

Internet Access

WiFi is available at the campus via eduroam and Plus_Event. Access data for Plus_Event is:

SSID: Plus_Event
Username: sim19
Password: F4zyh9us7g

Lunch

Lunch can be obtained at a very reasonable price at the mensa directly adjacent to the lecture halls. We have asked the mensa to provide some typical Austrian dishes during the week of our conference. There is always a vegetarian choice on the menu and additionally a salad bar. You may even sit outside (provided reasonable weather) and enjoy a nice view of the fortress while having lunch.

Coffee Breaks

Refreshments and coffee will be served during the coffee breaks in the foyer in front of the lecture room 402 (Blauer Hörsaal) and 403 (Grüner Hörsaal).

Reception

There will be a reception on Monday evening at 19:00 in the Salzburg "Residenz" Palace (Residenzplatz 1, 5020 Salzburg). A welcome address will be delivered by the Mayor of the City of Salzburg and the Governor of the State of Salzburg. If you plan on attending, please arrive on time.

Social Dinner

The social dinner will take place on Tuesday at 19:00 pm in the restaurant *Stiegl-Brauwelt* and starts with an aperitif. The dinner will start at about 19:45. The conference dinner is included in the conference fee. The address is:

Stiegl-Brauwelt
Bräuhausstraße 9
5020 Salzburg

There will be a bus transfer from the Faculty of Natural Science (NAWI) of the University to the restaurant at 18:20 pm and back at 10:00 pm and 11:00 pm.

Local Transportation

To reach the conference location, take the bus lines 3 or 8 (get off at the bus stop "Faistauergasse"). Bus 3 runs at least once every 8-10 minutes during daytime. For details and even ticket purchase, you may consider installing the Salzburg Verkehr App, but you can also buy tickets at a Trafik or from the bus driver. If you plan on using public transport frequently, a week ticket for € 17 may be a good choice. From the bus stop "Faistauergasse" walk straight ahead in the outbound direction, until you reach the next traffic light, then turn right onto "Michael-Pacher-Straße", and walk straight on until you reach the large yellow building.

Restaurants

There are many good restaurants in Salzburg. Therefore we can only list a small selection: (reservation is recommended)

Upmarket Austrian cuisine

Arthotel Blaue Gans
Getreidegasse 41–43
Phone: +43 662 84 24 91-0

K+K Restaurant am Waagplatz
Waagplatz 2
Phone: +43 662 84 21 56

Restaurant Wasserfall
Linzer Gasse 10
Phone: +43 662 87 33 31

International restaurants

Trattoria Domani
Kaigasse 33
Phone: +43 662 84 27 43

Lemonchilli
Nonntaler Haupstraße 24
Phone: +43 662 84 25 58

L'Osteria
Dreifaltigkeitsgasse 10
Phone: +43 662 87 06 58 10

Traditional Austrian cuisine

Augustinerbräu
Lindhofstraße 7
Phone: +43 662 43 12 46

Gablerbräu
Linzer Gasse 9
Phone: +43 662 88 965

Restaurant Stieglkeller
Festungsgasse 10
Phone: +43 662 84 26 81

Restaurant Triangel
Wiener Philharmonikergasse 7
Phone: +43 662 84 22 29

Zum fidelen Affen
Pristerhausgasse 8
Phone: +43 662 87 73 61

Cafés

There are several excellent Kaffeehäuser (Cafés) in Salzburg, and a long tradition of sophisticated coffee drinking. We are happy to let you know that Café Habakuk will honor your presence in their café and your consumption of their great caffeinated beverages by giving you one of their hand-made and rather delicious Mozartkugeln for *free*. Just show your name badge (and perhaps this page) to the waiter or waitress.

Café Habakuk, Familie Vonblon
Linzer Gasse 26, 5020 Salzburg
Phone: +43.662.874150
`http://www.cafe-habakuk.at`

Conference Agenda

10th International Workshop on Simulation and Statistics

Date: Monday, 02/Sep/2019

11:45am - 12:30pm	**Opening and Welcome Address** Location: **HS 401 (Auditorium Maximum)** Chair: **Arne Bathke**	
12:30pm - 1:30pm	**K1: Keynote Regina Liu** Location: **HS 401 (Auditorium Maximum)**	
1:30pm - 3:35pm	**ED1: Experimental Design** Location: **HS 402 (Blue Lecture Hall)** Chair: **Roberto FONTANA**	
1:30pm - 3:45pm	**DM1: Advances in Dependence Modelling and Copulas I** Location: **HS 403 (Green Lecture Hall)** Chair: **Wolfgang TRUTSCHNIG**	
3:35pm - 4:00pm	**Coffee break**	
4:00pm - 6:05pm	**DM2: Advances in Dependence Modelling and Copulas II** Location: **HS 403 (Green Lecture Hall)** Chair: **Wolfgang TRUTSCHNIG**	**SMS1: Stochastic Modelling and Simulation in Atmospheric Science** Location: **HS 402 (Blue Lecture Hall)** Chair: **Elena YAROVAYA**

Date: Tuesday, 03/Sep/2019

9:00am - 10:00am	**K2: Keynote Edgar Brunner** Location: **HS 401 (Auditorium Maximum)**		
10:00am - 10:30am	**Coffee break**		
10:30am - 12:10pm	**CMD1: Inference for Complex Multivariate Designs - Enhanced Procedures and open questions** Location: **HS 403 (Green Lecture Hall)** Chair: **Markus PAULY**	**ES1: Asymptotic Analysis and Simulation of Complex Stochastic Evolutionary Systems I** Location: **HS 402 (Blue Lecture Hall)** Chair: **Larisa AFANASEVA** Chair: **Ekaterina BULINSKAYA** Chair: **Elena YAROVAYA**	
12:10pm - 1:30pm	**Lunch**		
1:00pm - 1:30pm	**PS: Student Poster Session** Location: **Foyer HS 402 (Blue Lecture Hall)** Chair: **Cami Marie FUGLSBY** Chair: **Georg ZIMMERMANN**		
1:30pm - 3:10pm	**ES2: Asymptotic Analysis and Simulation of Complex Stochastic Evolutionary Systems II** Location: **HS 402 (Blue Lecture Hall)** Chair: **Ekaterina BULINSKAYA** Chair: **Larisa AFANASEVA** Chair: **Elena YAROVAYA**	**NE1: Statistical Challenges in Neurology** Location: **HS 403 (Green Lecture Hall)** Chair: **Yvonne HÖLLER**	
3:10pm - 4:25pm	**MT1: Recent Advances in Modelling Techniques and Hypothesis Testing Problems I** Location: **HS 402 (Blue Lecture Hall)** Chair: **Ilia VONTA**	**NP1: Nonparametric Inference I** Location: **HS 403 (Green Lecture Hall)** Chair: **Arne Bathke**	
4:25pm - 4:50pm	**Coffee break**		
4:50pm - 6:05pm	**FC1: Free contributions I** Location: **HS 403 (Green Lecture Hall)** Chair: **Judith PARKINSON**	**MT2: Recent Advances in Modelling Techniques and Hypothesis Testing Problems II** Location: **HS 401 (Auditorium Maximum)** Chair: **Ilia VONTA**	**OD1: Optimal Design of Experiments** Location: **HS 402 (Blue Lecture Hall)** Chair: **Victor CASERO-ALONSO**

Date: Wednesday, 04/Sep/2019

8:30am - 10:35am	**EDA1: Experimental Design and Application** Location: **HS 403 (Green Lecture Hall)** Chair: **Nina Marie SCHMITZBERGER**	**FS1: Forensic Statistics** Location: **HS 402 (Blue Lecture Hall)** Chair: **Christopher Paul SAUNDERS**	
10:35am - 11:00am	**Coffee break**		
11:00am - 12:40pm	**FD1: Functional Data** Location: **HS 403 (Green Lecture Hall)** Chair: **Alexander MEISTER**	**NP2: Nonparametric Inference II** Location: **HS 402 (Blue Lecture Hall)** Chair: **Rosa ARBORETTI**	**SL1: Statistical Learning - Methods and Applications I** Location: **HS 401 (Auditorium Maximum)** Chair: **Jürgen PILZ**
12:40pm - 1:30pm	**Lunch**		
1:30pm - 3:35pm	**SM1: Statistical Modelling and Data Analysis** Location: **HS 403 (Green Lecture Hall)** Chair: **Subir GHOSH**	**SU1: Recent Developments in Survival Analysis** Location: **HS 402 (Blue Lecture Hall)** Chair: **Marialuisa RESTAINO**	
3:35pm - 4:00pm	**Coffee break**		
4:00pm - 5:15pm	**ENS1: Environmental Statistics** Location: **HS 403 (Green Lecture Hall)** Chair: **Gunter SPÖCK**	**ESA1: Ergodicity and sensitivity analysis of stochastic systems and networks I** Location: **HS 401 (Auditorium Maximum)** Chair: **Elmira Yu. KALIMULINA**	**FTS1: Functional Time Series** Location: **HS 402 (Blue Lecture Hall)** Chair: **Siegfried HÖRMANN**
5:15pm - 6:30pm	**SL2: Statistical Learning - Methods and Applications II** Location: **HS 403 (Green Lecture Hall)** Chair: **Michael G. SCHIMEK**	**SSP1: Statistical Selection Procedures and Multiple Comparison** Location: **HS 402 (Blue Lecture Hall)** Chair: **Martin HAPP**	
6:30pm - 7:30pm	**K3: Keynote Gerd Antes** Location: **HS 401 (Auditorium Maximum)**		

Date: Thursday, 05/Sep/2019

9:00am - 10:00am	**K4: Keynote Christian Robert** Location: **HS 401 (Auditorium Maximum)** Chair: **Jürgen PILZ**		
10:00am - 10:30am	**Coffee break**		
10:30am - 12:10pm	**LF1: Likelihood-free Statistical Design and Inference** Location: **HS 403 (Green Lecture Hall)** Chair: **Werner G. MÜLLER**	**SSM1: Stochastic Modelling and Simulation in Materials Science and Engineering I** Location: **HS 402 (Blue Lecture Hall)** Chair: **Volker SCHMIDT**	
12:10pm - 1:00pm	**Lunch**		
1:00pm - 2:15pm	**CB1: Algebraic Methods in Computational Biology** Location: **HS 402 (Blue Lecture Hall)** Chair: **Ruriko YOSHIDA**	**MNT1: Modern Statistics and New Technologies** Location: **HS 403 (Green Lecture Hall)** Chair: **Michael G. SCHIMEK**	
2:15pm - 3:55pm	**BDA1: Big Data Analytics and High-Dimensional Data Analysis** Location: **HS 401 (Auditorium Maximum)** Chair: **S. Ejaz AHMED**	**DGA1: Data generation assisted inference** Location: **HS 402 (Blue Lecture Hall)** Chair: **Regina Y. LIU**	**SMR1: Statistical methods and applications in medical research** Location: **HS 403 (Green Lecture Hall)** Chair: **Georg ZIMMERMANN**

Date: Friday, 06/Sep/2019

9:00am - 10:00am	**K5: Keynote Holger Dette** Location: **HS 401 (Auditorium Maximum)**		
10:00am - 10:30am	**Coffee break**		
10:30am - 12:10pm	**FC2: Free contributions II** Location: **HS 403 (Green Lecture Hall)** Chair: **Martin HAPP**	**SL3: Statistical Learning - Methods and Applications III** Location: **HS 402 (Blue Lecture Hall)** Chair: **Jürgen PILZ** Chair: **Michael G. SCHIMEK**	**SSM2: Stochastic Modelling and Simulation in Materials Science and Engineering II** Location: **HS 401 (Auditorium Maximum)** Chair: **Volker SCHMIDT**
12:10pm - 1:00pm	**Lunch**		
1:00pm - 3:05pm	**OD2: Optimal Design in Mixed Models** Location: **HS 403 (Green Lecture Hall)** Chair: **Maryna PRUS**	**OSP1: Optimal Selection Procedures** Location: **HS 402 (Blue Lecture Hall)** Chair: **Dieter RASCH**	
3:05pm - 4:20pm	**AN1: Analytical and Numerical Methods in Statistics** Location: **HS 401 (Auditorium Maximum)** Chair: **Hans Dieter SCHOTT**	**ESA2: Ergodicity and sensitivity analysis of stochastic systems and networks II** Location: **HS 402 (Blue Lecture Hall)** Chair: **Elmira Yu. KALIMULINA**	**SNM1: Stochastic numerical methods** Location: **HS 403 (Green Lecture Hall)** Chair: **Svetlana Nikolaevna LEORA**

Abstracts

Keynote Speakers

Sessions

Monday

Tuesday

Location: HS 402 (Blue Lecture Hall)
Chair: Larisa AFANASEVA, Ekaterina BULINSKAYA, Elena YAROVAYA

ES2: Asymptotic Analysis and Simulation of Complex Stochastic Evolutionary Systems II (p. 31)
Location: HS 402 (Blue Lecture Hall)
Chair: Ekaterina BULINSKAYA, Larisa AFANASEVA, Elena YAROVAYA

NE1: Statistical Challenges in Neurology (p. 33)
Location: HS 403 (Green Lecture Hall)
Chair: Yvonne HÖLLER

MT1: Recent Advances in Modelling Techniques and Hypothesis Testing Problems I (p. 34)
Location: HS 402 (Blue Lecture Hall)
Chair: Ilia VONTA

NP1: Nonparametric Inference I (p. 36)
Location: HS 403 (Green Lecture Hall)
Chair: Arne Bathke

FC1: Free contributions I (p. 37)
Location: HS 403 (Green Lecture Hall)
Chair: Judith PARKINSON

MT2: Recent Advances in Modelling Techniques and Hypothesis Testing Problems II (p. 39)
Location: HS 401 (Auditorium Maximum)
Chair: Ilia VONTA

OD1: Optimal Design of Experiments (p. 40)
Location: HS 402 (Blue Lecture Hall)
Chair: Victor CASERO-ALONSO

Wednesday

EDA1: Experimental Design and Application (p. 42)
Location: HS 403 (Green Lecture Hall)
Chair: Nina Marie SCHMITZBERGER

FS1: Forensic Statistics (p. 44)
Location: HS 402 (Blue Lecture Hall)
Chair: Christopher Paul SAUNDERS

FD1: Functional Data (p. 46)
Location: HS 403 (Green Lecture Hall)
Chair: Alexander MEISTER

NP2: Nonparametric Inference II (p. 47)
Location: HS 402 (Blue Lecture Hall)
Chair: Rosa ARBORETTI

SL1: Statistical Learning - Methods and Applications I (p. 49)
Location: HS 401 (Auditorium Maximum)
Chair: Jürgen PILZ

SM1: Statistical Modelling and Data Analysis (p. 51)
Location: HS 403 (Green Lecture Hall)
Chair: Subir GHOSH

SU1: Recent Developments in Survival Analysis (p. 54)
Location: HS 402 (Blue Lecture Hall)
Chair: Marialuisa RESTAINO

ENS1: Environmental Statistics (p. 56)
Location: HS 403 (Green Lecture Hall)
Chair: Gunter SPÖCK

ESA1: Ergodicity and sensitivity analysis of stochastic systems and networks I (p. 58)
Location: HS 401 (Auditorium Maximum)
Chair: Elmira Yu. KALIMULINA

FTS1: Functional Time Series (p. 59)
Location: HS 402 (Blue Lecture Hall)
Chair: Siegfried HÖRMANN

SL2: Statistical Learning - Methods and Applications II (p. 60)
Location: HS 403 (Green Lecture Hall)
Chair: Michael G. SCHIMEK

SSP1: Statistical Selection Procedures and Multiple Comparison (p. 62)
Location: HS 402 (Blue Lecture Hall)
Chair: Martin HAPP

Thursday

LF1: Likelihood-free Statistical Design and Inference (p. 63)
Location: HS 403 (Green Lecture Hall)
Chair: Werner G. MÜLLER

SSM1: Stochastic Modelling and Simulation in Materials Science and Engineering I (p. 65)
Location: HS 402 (Blue Lecture Hall)
Chair: Volker SCHMIDT

CB1: Algebraic Methods in Computational Biology (p. 67)
Location: HS 402 (Blue Lecture Hall)
Chair: Ruriko YOSHIDA

MNT1: Modern Statistics and New Technologies (p. 68)
Location: HS 403 (Green Lecture Hall)
Chair: Michael G. SCHIMEK

BDA1: Big Data Analytics and High-Dimensional Data Analysis (p. 69)
Location: HS 401 (Auditorium Maximum)
Chair: S. Ejaz AHMED

DGA1: Data generation assisted inference (p. 71)
Location: HS 402 (Blue Lecture Hall)
Chair: Regina Y. LIU

SMR1: Statistical methods and applications in medical research (p. 73)
Location: HS 403 (Green Lecture Hall)
Chair: Georg ZIMMERMANN

Friday

FC2: Free contributions II (p. 75)
Location: HS 403 (Green Lecture Hall)
Chair: Martin HAPP

SL3: Statistical Learning - Methods and Applications III (p. 78)
Location: HS 402 (Blue Lecture Hall)
Chair: Jürgen PILZ, Michael G. SCHIMEK

SSM2: Stochastic Modelling and Simulation in Materials Science and Engineering II (p. 80)
Location: HS 401 (Auditorium Maximum)
Chair: Volker SCHMIDT

OD2: Optimal Design in Mixed Models (p. 82)
Location: HS 403 (Green Lecture Hall)
Chair: Maryna PRUS

OSP1: Optimal Selection Procedures (p. 84)
Location: HS 402 (Blue Lecture Hall)
Chair: Dieter RASCH

AN1: Analytical and Numerical Methods in Statistics (p. 85)
Location: HS 401 (Auditorium Maximum)
Chair: Hans Dieter SCHOTT

ESA2: Ergodicity and sensitivity analysis of stochastic systems and networks II (p. 86)
Location: HS 402 (Blue Lecture Hall)
Chair: Elmira Yu. KALIMULINA

SNM1: Stochastic numerical methods (p. 88)
Location: HS 403 (Green Lecture Hall)
Chair: Svetlana Nikolaevna LEORA

Poster Session

See page 89 for details.

Cami Marie FUGLSBY
Kiegan E RICE
Amy CRAWFORD
Patrick B. LANGTHALER
Bita SAMIMIZAD
Gert DEHNEN
Victoria RACHER

Keynote Speakers

Monday September 2, 2019

K1: Keynote Regina Liu
Auditorium Maximum (HS 401)

12:30pm - 1:30pm
Prediction With Confidence - General Framework for Predictive Inference
Presenter: Regina Y. LIU, RUTGERS UNIVERSITY, United States of America
We propose a general framework for prediction in which a prediction is in the form of a distribution function, called predictive distribution function. This predictive distribution function is well suited for prescribing the notion of confidence under the frequentist interpretation and providing meaningful answers for prediction-related questions. Its very form of a distribution function also lends itself as a useful tool for quantifying uncertainty in prediction. A general approach under this framework is formulated and illustrated using the so-called confidence distributions (CDs). This CD-based prediction approach inherits many desirable properties of CD, including its capacity to serve as a common platform for directly connecting the existing procedures of predictive inference in Bayesian, fiducial and frequentist paradigms. We discuss the theory underlying the CD-based predictive distribution and related efficiency and optimality. We also propose a simple yet broadly applicable Monte-Carlo algorithm for implementing the proposed approach. This concrete algorithm together with the proposed definition and associated theoretical development provide a comprehensive statistical inference framework for prediction. Finally, the approach is demonstrated by simulation studies and a real project on predicting the volume of application submissions to a government agency. The latter shows the applicability of the proposed approach to even dependent data settings.

Tuesday September 3, 2019

K2: Keynote Edgar Brunner
Auditorium Maximum (HS 401)

9:00am - 10:00am
Ranking Procedures In Unbalanced Factorial Designs - Surprising Results
Presenter: Edgar BRUNNER, University Medical School Göttingen, Germany
Rank-based inference methods are applied in various disciplines, typically when procedures relying on standard normal theory are not justifiable. Various specific rank-based methods have been developed for two and more samples, and also for general factorial designs. It is the aim of the present talk
1) to demonstrate that the common rank-procedures for several samples or general factorial designs may lead to seemingly paradoxical results in case of unequal sample sizes as compared to equal sample sizes,

2) to explain why this is the case, and
3) to provide a method to overcome these disadvantages.
Theoretical investigations show that the paradoxical results can be explained by considering the non-centralities of the test statistics which may be non-zero for the usual rank-based procedures in case of unequal sample sizes while they may be equal to 0 in case of equal sample sizes. A simple solution is to consider unweighted relative effects instead of weighted relative effects. The former effects are estimated by means of the so-called pseudo-ranks while the usual ranks naturally lead to the latter effects. The unweighted relative effects have a natural and intuitive interpretation and, moreover, they are fixed model quantities by which hypotheses can be formulated and for which confidence intervals can be derived. These procedures are based on pseudo-ranks and similar results as for the common rank procedures can be derived. The computations can be performed by the R-package rankFD which can be downloaded from CRAN. Handling of this package and the theoretical background are described in Brunner et al. (2019). A real data example illustrates the practical meaning of the theoretical discussions.
Reference
Brunner, E., Bathke, A.C., and Konietschke, F. (2019). Rank- and Pseudo-Rank Procedures for Independent Observations in Factorial Designs – Using R and SAS. Springer Series in Statistics, Springer, Heidelberg. ISBN: 978-3-030-02912-8.

Wednesday September 4, 2019

K3: Keynote Gerd Antes
Auditorium Maximum (HS 401)

6:30pm - 7:30pm
Evidence-based Medicine (EBM) and Big Data - Friends or Enemies?
Presenter: Gerd ANTES, Cochrane Germany
More than 25 years of Evidence-Based Medicine (EBM) have achieved major contributions to systematically integrating the results of clinical trials into decision making in health care. This progress has been enabled and supported by an enormous amount of methodological developments. The rigour of the methodological framework, in particular addressing quality assessment and quality in general, is a characteristic of EBM.
In recent years Big Data, artificial intelligence (AI) and personalized medicine have generated a realm of visions and promises where the quality issue seems to have completely disappeared: Unlimited data guarantee any level of needed quality, without particular effort. Can this be expected, or where is the border between realistic expectations and marketing-driven promotion?
We are observing a confrontation and a cultural clash between the “old”, methods-driven world and the new “informatics-based” world which is not receiving the attention it deserves in the current climate of enthusiasm and hype, to avoid misleading perspectives and return to strictly quality-driven research agendas and the implementation of these methods.

Thursday September 5, 2019

K4: Keynote Christian Robert
Auditorium Maximum (HS 401)

9:00am - 10:00am

Approximate Bayesian computation and inference

Presenter: Christian P ROBERT, University Paris Dauphine, France

The ABC algorithm appeared in the 1990s in complex genetic problems where the likelihood of the model is impossible to compute or even reliably approximate. The principle behind ABC is that, for a generative model, simulated data associated with a value of the model parameter can be compared with the true data and assess whether or not this parameter is likely to have generated the data. ABC methods are now standard tools in many branches of statistics when likelihood computation is an issue, including dynamic models in signal processing and financial data analysis, networks and queuing models. While these methods suffer from calibration difficulties that make their implementation delicate, a wide range of ABC versions has emerged, inspired from sequential Monte Carlo techniques as well as econometric methods, Bayesian nonparametrics, and learning tools such as random forests. In addition, ABC claims to validity include convergence as an estimation method and consistency for model choice, which represents a large part of its uses in applied domains. The lecture will covers both these validation steps and different implementations of ABC algorithms and calibration of their parameters.

Friday September 6, 2019

K5: Keynote Holger Dette
Auditorium Maximum (HS 401)

9:00am - 10:00am

Testing Relevant Hypotheses In Functional Time Series Via Self-normalization

Presenter: Holger DETTE, Ruhr-Universität Bochum, Germany

In this paper we develop methodology for testing relevant hypotheses in functional time series, but extensions to other settings are also discussed. Instead of testing for exact equality, for example for the equality of two mean functions from two independent time series, we propose to test a it relevant deviation under the null hypothesis. In the two sample problem this means that an L^2-distance between the two mean functions is smaller than a pre-specified threshold. For such hypotheses self-normalization, which was introduced by Shao (2010) and Shao and Zhang (2010) and is commonly used to avoid the estimation of nuisance parameters, is not directly applicable. We develop new self-normalized procedures for testing relevant hypotheses and demonstrate the particular advantages of this approach in the the comparisons of eigenfunctions of the covariance operator.

Sessions

Monday September 2, 2019

Opening and Welcome Address
Auditorium Maximum (HS 401)

11:45am - 12:00pm
Welcome Address
Presenter: Arne BATHKE, University of Salzburg, Austria
Welcome address Local Organizing Committee

12:00pm - 12:30pm
Ten Workshops on Simulation and Statistics
Presenter: Viacheslav Borisovich MELAS, St. Petersburg State University, Russian Federation
A brief history of the conference series "Workshops on Simulation".

ED1: Experimental Design
Blue Lecture Hall (HS 402)

1:30pm - 1:55pm
Echelon designs, Hilbert series and Smolyak grids
Presenter: Hugo MARURI AGUILAR, Queen Mary University of London
Co-Authors: Henry WYNN
Echelon designs were first described in the monograph by Pistone et al. (2000). These designs are defined for continuous factors and include, amongst others, factorial designs. They have the appealing property that the saturated polynomial model associated to it mirrors the geometric configuration of the design. Perhaps surprisingly, the interpolators for such designs are based upon the Hilbert series of the monomial ideal associated with the polynomial model and thus the interpolators satisfy properties of inclusion-exclusion. Echelon designs are quite flexible for modelling and include the recently developed designs known as Smolyak sparse grids. In our tal we present the designs, describe their properties and show examples of application.
This is joint work with H. Wynn (CATS, LSE).
Reference: Pistone et al. (2000) Algebraic Statistics. Chapman & Hall/CRC

1:55pm - 2:20pm
Representation Of Multivariate Bernoulli Distributions With a Given Set Of Sepecified Moments
Presenter: Patrizia SEMERARO, Politecnico di Torino, Italy

Co-Authors: Roberto FONTANA
We propose a new but simple method to characterise multivariate Bernoulli variables belonging to a given class, i.e., with some specified moments.
Within a given class, this characterisation allows us to easily generate a sample of mass functions.
It also provides the bounds that all the moments must satisfy to be compatible and the possibility to choose the best distribution according to a certain criterion.
For the special case of the Fr'echet class of the multivariate Bernoulli distributions with given margins we find a polynomial characterization of the class. Our characterization allows us to have bounds for the higher order moments. An algorithm and its use in some examples is shown. Possible connections with design of experiments will be shortly illustrated.

2:20pm - 2:45pm
Sequential Monte Carlo For Fredholm Integral Equations Of The First Kind
Presenter: Francesca Romana CRUCINIO, University of Warwick, United Kingdom
Co-Authors: Adam Michael JOHANSEN, Arnaud DOUCET
Fredholm integral equations of the first kind

$$h(y) = \int g(y \mid x) f(x) dx \tag{1}$$

describe a wide set of problems in science (e.g. image processing for motion deblurring and positron emission tomography) where the output data distribution h is a distorted version of the input signal f. A popular method to approximate f is an infinite dimensional Expectation-Maximization (EM) algorithm that, given an initial guess for f, iteratively refines the approximation by including the information given by h and g. We use Sequential Monte Carlo (SMC) to develop a stochastic discretisation of the Expectation-Maximization-Smoothing (EMS) algorithm, a regularised variant of EM. This stochastic discretisation can be implemented when only samples from h are available and g can be evaluated pointwise. We show that the approximations given by the resulting SMC algorithms converge to the solution of (1) in the weak topology and we show that the proposed approach outperforms the widely used discretisation by binning of the EMS in terms of accuracy of the reconstruction of f on 1-dimensional continuous mixtures.

2:45pm - 3:10pm
Optimal Paired Comparison Designs For Second-Order Interactions With Profile Strength Constraint
Presenter: Eric NYARKO, Otto-von-Guericke-University Magdeburg, Germany
Co-Authors: Rainer SCHWABE
In many fields of applications like marketing, psychology, transport economics and health economics often paired comparisons involving only options specified by a subset of the attributes (partial-profiles) to mitigate cognitive burden are used.

For the corresponding situation of linear paired comparisons, when continuous response is available for the amount of preferences, the problem of finding optimal designs is considered in the presence of second-order interactions. The resulting designs are also optimal for the situation of binary response in discrete choice under the indifference assumption of equal choice probabilities.

DM1: Advances in Dependence Modelling and Copulas I
Green Lecture Hall (HS 403)

1:30pm - 1:50pm
Copula-based Estimation Of The Cross Ratio Function
Presenter: Noel VERAVERBEKE, University of Hasselt
For a pair (T_1, T_2) of absolutely continuous random variables, the cross ratio function is defined as $\delta(t_1, t_2) = \lambda(t_1|T_2 = t_2)/\lambda(t_1|T_2 > t_2)$, where $\lambda(t_1|T_2 = t_2)$ and $\lambda(t_1|T_2 > t_2)$ are the conditional hazard rate functions of T_1 , given $T_2 = t_2$ and $T_2 > t_2$ respectively. Independence between T_1 and T_2 corresponds to$\delta(t_1, t_2) = 1$ and positive association corresponds to $\delta(t1, t2) > 1$. Nowadays the cross ratio function is a commonly used measure to describe local dependence between two correlated random variables.
Being a ratio of conditional hazard functions, the cross ratio can be written in terms of the survival copula of T1 and T2 and its partial derivatives. Using Bernstein estimators for the survival copula and its derivatives, we obtain Bernstein based estimators for the conditional hazards and a nonparametric estimator for the cross ratio function $\delta(t_1, t_2)$. The reason for using a Bernstein copula-based estimator for the cross ratio function is motivated from earlier results showing good bias and variance properties. The asymptotic distributional behavior of the new estimator is established.
We also consider a number of simulations to study the finite sample performance for copulas with different types of local dependence. A real data set on asthma attacks in children is used to investigate the local dependence between event times in the placebo and treated groups.

1:50pm - 2:15pm
On Kendall's Tau for Order Statistics
Presenter: Sebastian FUCHS, TU Dortmund, Germany
Co-Authors: Klaus D. SCHMIDT
Every copula C for a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with identically distributed coordinates determines a unique copula $C_{:d}$ for its order statistic $\mathbf{X}_{:d} = (X_{1:d}, \ldots, X_{d:d})$. We study the dependence structure of $C_{:d}$ via Kendall's tau, denoted by κ. As a general result, we show that $\kappa(C_{:d})$ is at least as large as $\kappa(C)$. For the product copula Pi, which corresponds to the case of independent coordinates of $\mathbf{X}$, we provide an explicit formula for $\kappa(Pi_{:d})$ showing that the inequality between $\kappa(\Pi)$ and $\kappa(Pi_{:d})$ is strict. We also compute Kendall's tau for certain multivariate margins of $\Pi_{:d}$ corresponding to the lower or upper coordinates of $\mathbf{X}_{:d}$.

2:15pm - 2:35pm

On a Construction of Multivariate Distributions Given Some Multidimensional Marginals

Presenter: Didier RULLIÈRE, University Lyon

Co-Authors: Nabil KAZI-TANI

We investigate the link between the joint law of a d-dimensional random vector and the law of some of its multivariate marginals. We introduce and focus on a class of distributions, that we call projective, for which we give detailed properties. This allows us to obtain necessary conditions for a given construction to be projective. We illustrate our results by proposing some theoretical projective distributions, as elliptical distributions or a new class of distribution having given bivariate margins. In the case where the data do not necessarily correspond to a projective distribution, we also explain how to build proper distributions while checking that the distance to the prescribed projections is small enough.

2:35pm - 3:00pm

Construction, Sampling and Estimation of Hierarchical Outer Power Archimedean copulas

Presenter: Jan GÓRECKI, Silesian University in Opava, Czech Republic

Co-Authors: Marius HOFERT, Ostap OKHRIN

Distributions based on hierarchical Archimedean copulas (HACs) became popular as they enable one to model non-elliptical and non-exchangeable dependencies among random variables. Their practical applications reported in the literature are, however, mostly limited to the case in which all generator functions in a HAC are one-parametric, which implies that all properties (e.g., Kendall's tau and tail dependence coefficients) of each bivariate margin of such a HAC is given just by a single parameter. Involving so-called outer power transformations of Archimedean generators in such models, this limitation can be alleviated, which typically allows one to set Kendall's tau and upper-tail dependence coefficient independently of each other. This talk addresses the construction, sampling and estimation of the resulting so-called hierarchical outer power Archimedean copulas.

3:00pm - 3:20pm

On Some Properties of Reflected Maxmin Copulas

Presenter: Noppadon KAMNITUI, Department of Mathematics, Paris-Lodron-University of Salzburg

Co-Authors: Wolfgang TRUTSCHNIG

The class of maxmin copulas extends the well known classes of Marshall- Okin and Marshall copulas by allowing the external shocks to have different effects on the components. To study the maxmin copula one introduced the class of so-called reflected maxmin copulas (RMM) by using a reflection in one of the variables. We show that the family of all RMM copulas is compact subset of the metric space of all copulas, characterize absolutely continuous RMM copulas, and the limit of absolutely continuous RMM copulas is

absolutely continuous too. Finally, we determine Kendall's Tau and Spearman's Rho of RMM copulas and prove some inequalities for Tau and Rho .

3:20pm - 3:45pm
Zero-sets of copulas
Presenter: Enrique DE AMO ARTERO, Universidad de Almería, Spain
We study conditions on sets in order to be zero-sets of semi- copulas, quasi-copulas, and with special attention, of copulas. We find necessary and sufficient conditions for the zero-sets of absolutely continuous copulas and copulas whose support coincides with the closure of the complementary to the zero-set. Moreover, we study several topological properties and the lattice-theoretic structure, and characterize the zero-sets of the class of Archimedean copulas.
This is joint work with J. Fernández Sánchez and M. Úbeda Flores.

DM1: Advances in Dependence Modelling and Copulas II
Green Lecture Hall (HS 403)

4:00pm - 4:25pm
Analysing the Relationship between District Heating Demand and Weather Conditions through Conditional Copula
Presenter: F. Marta L. DI LASCIO, Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
Co-Authors: Andrea MENAPACE, Maurizio RIGHETTI
We perform a copula-based analysis of thermal energy demand and its complex relationship with meteorological variables, such as solar radiation and temperature. The gathered data concern the district heating system (DHS) of the Italian city Bozen-Bolzano and the weather station S. Maurizio during 2014-2017. Since modern DHSs are sustainable energy distribution services that exploit renewable sources and avoid energy waste, in-depth knowledge of thermal energy demand, which is mainly affected by weather conditions, is essential to enhance heat production schedules and contribute to reducing climate change. Hence, we perform a three-step analysis that consists in i) removing serial dependence in each time series using autoregressive integrated moving average models, ii) estimating copula models on residual time series, and iii) analytically deriving the conditional copula-based probability function of thermal energy demand given meteorological variables. The investigation of the probability law of energy demand given weather scenarios, especially extreme climatic phenomena, makes it possible to provide useful insight on the production management phase of local energy utilities.

4:25pm - 4:50pm
Geometric Structure in Dependence Models and Applications
Presenter: Elisa PERRONE, Massachusetts Institute of Technology, United States of America

The growing availability of data makes it challenging yet crucial to model complex dependence traits. For example, hydrological and financial data typically display tail dependences, non-exchangeability, or stochastic monotonicity. Copulas serve as tools for capturing these complex traits and constructing accurate dependence models which resemble the underlying distributions of data. This talk explores the geometric properties of copulas to address dependence modeling challenges in several applications, such as hydrology and finance. In particular, we study the class of discrete copulas, i.e., restrictions of copulas on uniform grid domains, which admits representations as convex polytopes. In the first part of the talk, we give a geometric characterization of discrete copulas with desirable stochastic constraints in terms of the properties of their associated convex polytopes. In doing so, we draw connections to the popular Birkhoff polytopes, thereby unifying and extending results from both the statistics and the discrete geometry literature. In the second part of the talk, we further consolidate the statistics/discrete geometry bridge by showing the significance of our geometric findings to (1) construct entropy-copula models useful in hydrology, and (2) design test statistics for stochastic monotonicity properties of interest in finance.

4:50pm - 5:15pm
Clustering Ranking Data Via Copulas
Presenter: Marta NAI RUSCONE, LIUC, Italy
Clustering of ranking data aims at the identification of groups of subjects with a homogenous, common, preference behavior. Ranking data occurs when a number of subjects are asked to rank a list of objects according to their personal preference order. The input in cluster analysis is a distance matrix, whose elements measure the distances between rankings of two subjects. The choice of the distance dramatically affects the final result and therefore the computation of an appropriate distance matrix is an issue. Several distance measures have been proposed for ranking data. The most important are the Kendall's t, Spearman's r and Cayley distances. When the aim is to emphasize top ranks, weighted distances for ranking data should be used. We propose a generalization of this kind of distances using copulas. Those generalizations provide a more flexible instrument to model different types of data dependence structures and consider different situationsin the classification process. Simulated and real data are used to illustrate the pertinence and the importance of our proposal.

5:15pm - 5:40pm
Quantifying And Estimating Asymmetric Dependence
Presenter: Wolfgang TRUTSCHNIG, University of Salzburg, Austria
Co-Authors: Florian GRIESSENBERGER, Robert R. JUNKER
Standard dependence measures considered in the (mostly non-mathematical) literature like Pearson correlation, Spearman correlation, the Maximal information coefficient (MIC), and Schweitzer and Wolff's famous sigma are symmetric, i.e. they assign each pair (X, Y) of random variables the same dependence as they assign the pair (Y,X). Independence

of two random variables is a symmetric concept modelling the situation that knowing X does not change our knowledge about Y and vice versa - dependence, however, is not. Thinking, for instance, of a sample $(x_1, y_1), ..., (x_n, y_n)$ roughly in the shape of a noisy letter V, it is without doubt (on average) easier to predict the y-value given the x-value than vice versa. The R-package qad (short for quantification of asymmetric dependence) aims at detecting asymmetries in samples. It estimates the dependence of the second variable on the first one and vice versa, and additionally quantifies the asymmetry of the underlying dependence structure. The main objectives of the talk are to sketch the basic ideas behind qad, to present the most relevant mathematical properties of the underlying estimator(s), and to illustrate its capabilities by some examples.

5:40pm - 6:05pm
Qad: An R-Package For Quantifying (Asymmetric) Dependence
Presenter: Florian GRIESSENBERGER, Paris-Lodron-University of Salzburg
Co-Authors: Robert JUNKER, Wolfgang TRUTSCHNIG
Standard dependence measures like Pearson correlation, Spearman correlation, and Schweitzer and Wolff's σ are symmetric, i.e. they assign the pair (X, Y) of random variables X, Y the same dependence as the pair (Y, X). Since general dependence structures can be highly asymmetric, we developed a strongly consistent estimator of a copula-based, asymmetric dependence measure ζ_1 introduced in 2010, by using so-called empirical checkerboard copulas and Markov kernels. We implemented the estimator and related functions (quantifying asymmetry, testing for equal dependence, etc.) in the R-package qad available on CRAN. In the talk we introduce the R-package qad, illustrate the main features like consistency of the estimator and present some first results for the discrete setting.

SMS1: Stochastic Modelling and Simulation in Atmospheric Science
Blue Lecture Hall (HS 402)

4:00pm - 4:25pm
Weather Generators And Time Varying Dependence For Long-term Drought Risk
Presenter: Ozan EVKAYA, Atilim University, Turkey
For a specific sub-region, it is important to highlight the possible impacts of future droughts to cope with the variability of extreme events. In that respect, the effect of climate variability is studied extensively but the results are limited by lack of long term continuous climate data. To fix this problem, numerous stochastic weather generators are proposed to simulate realistic atmospheric variables like rainfall easily and these models provide suitable realizations for risk assessment in the design of water resource and environmental systems. In this case study, the long term climate variability will be studied using various multi-site weather generators. After the construction of synthetic climate variables, a detailed drought analysis will be investigated in terms of widely used drought index incorporating copula framework. Not last but not least, the findings of the study will be coupled with dependence analysis to discuss weather based insurance market which

can protect farmers against crop losses in Turkey.

4:25pm - 4:50pm

Numerical Stochastic Models of Conditional Non-Gaussian Processes with Conditions in the Form of Inequalities for Solving Problems of Statistical Meteorology

Presenter: Nina KARGAPOLOVA, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russian Federation; Novosibirsk State University, Russian Federation
Co-Authors: Vasily OGORODNIKOV

In this paper an algorithm for numerical simulation of conditional non-Gaussian processes with conditions in the form of inequalities is proposed. Analytical expressions for conditional distribution functions and distribution densities are obtained under the condition that values of a random process belong to the specified intervals. Results of a comparison of this algorithm with an algorithm based on the method of full enumeration under various sets of conditions are presented. A modification of the proposed algorithm for the case of simulating of a conditional random sequence in which at each simulation step the condition is imposed only on the process value preceding the simulated one is also presented. Examples of using the considered algorithms for simulation of the non-stationary conditional time series of air temperature are given.
This work was partly financially supported by the Russian Foundation for Basic Research (grant No 18-01-00149-a).

4:50pm - 5:15pm

Stochastic Models of Joint Non-stationary Meteorological Time Series to Study the Bioclimatic Indices

Presenter: Vasily OGORODNIKOV, Novosibirsk State University, Russian Federation
Co-Authors: Marina AKENTEVA, Nina KARGAPOLOVA, Elena KHLEBNIKOVA

In this paper a numerical stochastic model of the joint time-series of air temperature, relative humidity and wind speed modulus is proposed. The model takes into account the diurnal and seasonal variation of the real meteorological processes. Simulated trajectories of the joint time series are used to study various statistical properties of the time series of several bioclimatic indices that characterize the thermal effects of cold air and wind on human beings. The results of the study of the dependence of the properties of the bioclimatic indices on changes in the properties of the meteorological time series under consideration are presented.
This work was partly financially supported by the Russian Foundation for Basic Research (grant No 18-01-00149-a).

5:15pm - 5:40pm

An Adaptive Modelling Algorithm In The Problem Of Scattering Matrix Reconstruction

Presenter: Anna KORDA, The Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia
Co-Authors: Sergey UKHINOV

Two Monte Carlo weighted vector algorithms for statistical modelling of polarized radiation transfer are considered: a standart one, in which scattering angle is modeled by the first element of scattering matrix, and an adaptive one, in which scattering angle distribution density in every collision is chosen so as to ensure the finite variance of estimate. In this work algorithms for solving the problem of reconstructing the aerosol scattering matrix from ground-based observations of radiation in almukantarat of the Sun are proposed that use an adaptive method for simulating scattering in the atmosphere. The effectiveness of these methods in the "predictor-corrector" method of reconstructing the first two components of the scattering matrix is investigated by methods of numerical statistical modeling.

The reported study was partially funded by RFBR according to the research projects 17-01-00823, 18-31-00213, 18-01-00356.

5:40pm - 6:05pm

On Two Approaches to Estimating the Bidirectional Angular Densities of Distribution of Polarized Radiation

Presenter: Natalya V. TRACHEVA, Mechanics and Mathematics Department, Novosibirsk State University, Russia
Co-Authors: Sergey A. UKHINOV

In this talk, we discuss two different approaches to the numerical solution of one particular problem of atmospheric optics - the approximation of bidirectional angular characteristics of the polarized radiation, transmitted and back-scattered by the absorbing and scattering layer. The first approach is based on the orthonormal with Lambertian weight polynomial expansion of the bidirectional angular probability density function. The possibility of applying the Monte Carlo method is determined by the fact that expansion coefficients are the mathematical expectations of weighted random values of the standard functions. The second approach is a two-dimensional kernel density estimator. We discuss a specific for the posed problem technique of choosing optimal kernel bandwidth for uniform kernel function. For both approaches, we provide a comparative review of numerical results, obtained for angular probability densities of the intensity and the degree of polarization of radiation.

The reported study was partially funded by RFBR according to the research projects 17-01-00823, 18-31-00213, 18-01-00356.

Tuesday September 3, 2019

CMD1: Inference for Complex Multivariate Designs - Enhanced Procedures and open questions
Green Lecture Hall (HS 403)

10:30am - 10:55am

Testing Contrasts Of Quantiles In General Factorial Designs

Presenter: Marc DITZHAUS, TU Dortmund, Germany

Co-Authors: Markus PAULY

We consider the simultaneous estimation of various quantiles from different groups and derive multivariate central limit theorems in this context. Applying a Wald-type statistic to the resulting estimators leads to asymptotically exact tests in factorial designs. To obtain a better finite sample performance, we suggest a resampling version having the same preferable asymptotic properties as the asymptotic Wald-test under the null as well as under alternatives. As special cases, we obtain tests for the median in general factorial designs as well as k-sample tests for the interquartile range. Simulations analyze the small sample properties and the procedure's applicability is demonstrated in a real data example.

10:55am - 11:20am

General MANOVA with Missing Data - A Resampling-based Solution

Presenter: Lubna AMRO, Technical University of Dortmund, Germany

Co-Authors: Burim RAMOSAJ, Markus PAULY

Repeated measure designs and split plot plans are widely employed in scientific and medical research. The analysis of such designs is typically based on MANOVA models, requiring complete data, and certain assumption on the underlying parametric distribution such as normality or covariance matrix homogeneity. Several nonparametric multivariate methods have been proposed in the literature. They overcome the distributional assumptions, but the issue with missing data remains. The aim of this work is to develop asymptotic correct procedures that are capable of handling missing values without assuming normality, and allowing for covariance matrices that are heterogeneous between groups. This is achieved by applying proper resampling methods. The asymptotic theory for our suggested approaches is methodologically validated. Their small sample performance is further studied in an extensive simulation study and compared to existing methods such as multiple imputation based MANOVA. Finally, an illustrative data example is analyzed

11:20am - 11:45am

Testing Hypotheses about Covariance Matrices in General MANOVA Designs

Presenter: Paavo SATTLER, TU Dortmund, Germany

Co-Authors: Markus PAULY, Arne BATHKE

While testing mean-based hypotheses was mostly in the focus the last years, sta- tistical methods for inference on covariance matrices are as well of substantial interest. There is a need for hypothesis tests on covariance matrices not only as a preliminary test but also an independent test. Existing tests are either merely for specialized situations, such as testing equality of covariance matrices or have strict distributional assumptions like multivariate normality or elliptical distribu- tions. The goal of this work is to introduce a very

general approach to statistical hypoth- esis testing where the hypotheses are formulated in terms of covariance matrices. This includes, for example, hypotheses formulated using their traces, hypotheses of equality, and hypotheses for a given covariance matrix. The test procedures are based on two resampling approaches whose asymptotic validity are shown theoretically, while the actual finite sample performances have been investigated by means of extensive simulation studies.

11:45am - 12:10pm

Uncertainty Estimation with Bayesian Convolutional Neural Networks and Variational Inference

Presenter: Felix LAUMANN, Imperial College London, United Kingdom
Co-Authors: Kumar SHRIDHAR

We introduce a novel uncertainty estimation for classification tasks for Bayesian convolutional neural networks with variational inference. By normalizing the output of a Softplus function in the final layer, we estimate aleatoric and epistemic uncertainty in a coherent manner. The intractable posterior probability distributions over weights are inferred by Bayes by Backprop. Firstly, we demonstrate how this reliable variational inference method can serve as a fundamental construct for various network architectures. On multiple datasets in supervised learning settings (MNIST, CIFAR-10, CIFAR-100), this variational inference method achieves performances equivalent to frequentist inference in identical architectures, while the two desiderata, a measure for uncertainty and regularization are incorporated naturally. Secondly, we examine how our proposed measure for aleatoric and epistemic uncertainties is derived and validate it on the aforementioned datasets.

ES1: Asymptotic Analysis and Simulation of Complex Stochastic Evolutionary Systems I
Blue Lecture Hall (HS 402)

10:55am - 11:20am

The Formation of Particle Clusters in Branching Random Walks on Lattices

Presenter: Elena YAROVAYA, Lomonosov Moscow State University
Co-Authors: Daria BALASHOVA, Stanislav MOLCHANOV

Stochastic evolutionary systems with generation and transport of particles on the lattice, are usually called branching random walks. It is convenient to describe such processes in terms of birth, death, and walks of particles on the lattice. Such processes are used in numerous applications, in particular, in population dynamics. We consider a continuous-time symmetric branching random walk on the d-dimensional lattice, $d > 1$, and assume that at the initial moment there is one particle at every lattice point. Moreover, we assume that the underlying random walk has a finite variance of jumps and the reproduction law is described by a critical Bienamye-Galton-Watson process at every lattice point. We study the structure of the particle subpopulation generated by the initial particle situated at a lattice point x. We answer why vanishing of the majority of subpopulations leads to clusterization for the cases when the underlying random walk is recurrent, i.e. for lattice

dimensions d=1 and d=2. In the last decade, for various models of branching random walks, a series theoretical results were obtained among which the limit theorems about behavior of the process for large times has been obtained. However, as it is often the case, much more difficult or almost impossible to analyze analytically branching random walks on finite time intervals. Some necessary theoretical results for an interpretation of the simulation are given. The approach developed in this study is in good agreement with direct numerical simulations.
D. Balashova and E. Yarovaya were supported by the Russian Foundation for Basic Research (RFBR), project No. 17-01-00468. S. Molchanov was supported by the Russian Science Foundation (RSF), project No. 17-11-01098.

11:20am - 11:45am
Stochastic Modeling of Systems of Communicating Populations
Presenter: Manuel MOTA, University of Extremadura, Spain
Co-Authors: Manuel MOLINA, Casimiro CORBACHO
Stochastic processes provide a powerful tool for the modeling of populations and the posterior analysis of their asymptotic behaviour. The models based on stochastic processes are flexible enough to describe migration events in such populations.
In this work, we propose a new model to describe the simultaneous evolution of several populations which present migration flows among them. The model can be viewed as a modification of a multitype Galton-Watson branching process. We show also an application of this model to study the dynamics of several populations of big raptors, each of them located not very far from the others.

11:45am - 12:10pm
Numerical Simulation of the Interference for Wireless Device-to-Device Communications in Subway Trains Using the Non-Stationary Fokker-Planck Equation
Presenter: Yuliya GAIDAMAKA, Peoples' Friendship University of Russia (RUDN University), Federal Research Center "Computer Science and Control" of RAS, Russian Federation
Co-Authors: Yury ORLOV, Konstantin SAMOUYLOV
In this paper, the numerical simulation of the distribution function of the signal-to-interference ratio (SIR) in wireless device-to-device communication (D2D) is analyzed using the kinetic equation of the Fokker-Planck equation for the case of moving receivers and transmitters. This characteristic is a non-linear functional of the distance between the receiver and the transmitters, which transmit both the useful and the interfering signal. The stationarity of the distribution function of the signal-to-interference ratio depending on the motion parameters of moving receiving-transmitting devices of subscribers is investigated in the paper. This is an important characteristic that affects the stability of the connection — when the signal-to-interference ratio drops below a certain level, the connection is interrupted. In this paper, the problem of wireless D2D communication was

applied to subscribers in subway cars, the movement of which is regular, albeit random. It has been found that although the signal-to-interference ratio has a stochastic nonstationary behavior, the outage periods may have a regular structure. It is shown that the normalized average value (the so-called Sharpe coefficient) can still be considered as an indicator of the stability of the wireless connection. Thus, the method of stochastic control can use this indicator as the main control parameter, more, minimizing dispersion is more efficient than increasing the average SIR.
The publication has been prepared with the support of the "RUDN University Program 5-100" and funded by RFBR according to the research projects No. 17-07-00845, 18-07-00576

ES2: Asymptotic Analysis and Simulation of Complex Stochastic Evolutionary Systems II
Blue Lecture Hall (HS 402)

1:30pm - 1:55pm
Insurance Models with Dividends and Investments
Presenter: Ekaterina BULINSKAYA, Lomonosov Moscow State University, Russian Federation
Co-Authors: Boris SHIGIDA
The main goal of any insurance company is indemnification of the policyholders. Hence a company is interested in growth of its capital and tries to make the profitable investments. On the other hand, being a corporation, a company has to pay dividends to its shareholders. So, modeling an insurance company performance, it is necessary to combine reliability and cost approaches. In particular, treating the dividends, one aims at maximization of expected dividends paid out before the company ruin. To this end one can use such tools as investment, reinsurance or bank loans. Consideration of delays in bankruptcy procedure or dividends payment became also very popular in the last decade. All these aspects are taken into account in our study. We analyze several continuous-time and discrete-time models establishing optimal (and/or asymptotically optimal) policies. Discrete-time models turned out to be more appropriate in certain situations. Thus, dividends payments are usually effectuated on the base of the financial year results. The same is true with reinsurance, a treaty is negotiated by the end of a year as well. Numerical results are provided for illustration of theoretical ones. For simulation purpose it is important to establish the most significant parameters in the model description. Hence, we carry out the sensitivity analysis of the model to small fluctuations of parameters and perturbations of the underlying processes. Another problem, important for applications, is parameters estimation. It is also treated in presentation. The research is partially supported by the Russian Foundation for Basic Research under grant 17-01-00468.

1:55pm - 2:20pm
Sensitivity Analysis Of k-out-of-n System Characteristics To Shapes of Their Components Life and Repair Times Distributions
Presenter: Nika Mikhailovna IVANOVA, RUDN-University, Russian Federation

Co-Authors: Vladimir Vasilevich RYKOV
The paper deals with the problem about sensitivity of k-out-of-n system reliability characteristics to the shape of their elements life and repair times distribution under "quick" restoration. This research field has wide application in various fields, so its study dates back to the middle of the last century. In some previous works, closed form representations for reliability characteristics of hot redundant systems with exponential distribution of their elements life time and general distribution of their repair time have been found. The results allow to show their asymptotic insensitivity to the shape life and repair time distributions. In the recent paper sensitivity analysis of k-out-of-n system was carried out with the help of simulation modeling.

2:20pm - 2:45pm
Stability Conditions For Queueing Systems With Simultaneous Independent Service
Presenter: Svetlana GRISHUNINA, Lomonosov Moscow State University, Russian Federation; Moscow Institute of Electronics and Mathematics, National Research University Higher School of Economics, Russian Federation
In this paper we study the stability conditions of the systems with m identical servers in which customers arrive according to a regenerative input flow X(t). An arrived customer requires service from a random number of servers simultaneously.
A customer who arrives when the queue is empty begins service immediately when the number of servers he requires is available. If a customer arrives to the system when the queue is not empty he goes to the end of the queue. We consider systems with independent service where service times by different occupied servers of a given customer are independent.
We compare stability conditions for the considered queueing systems for different number of servers and different service times distribution and perform a numerical analysis of dependence of the stability conditions upon service discipline and distribution of service times.
Work is partially supported by Russian Foundation for Basic Research grant 17-01-00468.

2:45pm - 3:10pm
Stability Conditions And Statistical Analysis Of Retrial Queueing Systems
Presenter: Larisa AFANASEVA, Lomonosov Moscow State University, Russian Federation
Co-Authors: Elena BASHTOVA
We consider a multi-server queueing system with a regenerative input flow. An arriving customer finding one or more servers idle obtains service immediately. Customers who find all servers busy go directly to the orbit from which repeat attempts to get into idle server. In the classical retrial policy the intervals between successive repeated attempts are exponentially distributed with rate $\nu(j) = j\nu$, when there are j customers on the orbit. We consider more general models in which $\nu(j)$ is monotone increasing function

and refer these models to the first class. The second class contains models with constant retrial rate. This retrial policy is a useful device for modeling the retrial phenomenon in communication and computer networks. Based on the synchronization of the input flow and an auxiliary service process we establish the necessary and sufficient stability conditions for the models of the both classes. Furthermore, a consistent estimate of the system load coefficient is proposed.

NE1: Statistical Challenges in Neurology
Green Lecture Hall (HS 403)

1:30pm - 1:55pm

Representational Similarity Analyses Applied to Human Single Neuron Data

Presenter: Thomas P REBER, University of Bonn, Germany
Co-Authors: Florian MORMANN

A rather recent method to investigate the nature of information encoded in brain activity is representational similarity analysis (RSA). RSA can be applied to all types of multivariate brain activity measures and has been instrumental to further elucidate the complex relationships between neuronal activity and cognitive phenomena such as perception and memory. RSAs are based on matrices denoting measures of all pairwise distances between single-neuron patterns of brain activity evoked by cognitive states of interest (e.g., perceptions of different images). In this talk, I will introduce how such dissimilarity matrices (DM) can be computed and discuss a statistical significance test based on a label-shuffling procedure suited for testing hypotheses concerning the structure in DMs. I will discuss these methods in the context of a dataset of human single-neuron data obtained from the medial temporal lobes (MTL) of epilepsy patients undergoing chronic seizure monitoring. These data were recorded while participants viewed a set of images that can be grouped according to semantic features on multiple levels of abstraction (e.g., labeling an image as either 'terrier', 'dog', or 'animal'). Here, the multivariate RSA approach elucidates that information in the human MTL is encoded on a higher level of semantic abstraction than suggested by previous, univariate analyses of response profiles of individual neurons.

1:55pm - 2:20pm

Practical Considerations Related To The Detection Of Rare Events In The Human Electrencephalogram

Presenter: Matthias DÜMPELMANN, Epilepsy Center, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

Current treatment concepts for epilepsy are based on a continuous drug delivery or electrical stimulation to prevent the occurrence of seizures, exposing the brain and body to mostly unneeded risk of adverse effects. Responsive neurostimulation devices using state of the art signal processing and machine learning approaches promise to aim at timely interventions. But, seizures being rare and comparable short events challenge the training of machine learning algorithms. Here an overview about state of the art training environments, feature based classifiers and convolutional neural networks for seizure detection is given. It can be shown, that taken boundary conditions of low power hardware

into account a realization of feature-based classifiers and convolutional neural networks is possible in devices suited for implants. Finally an outlook is given how the generation of artificial training data and the analysis of long term-trends in seizure occurrence statistics has the potential to lead to substantial device improvement.

2:20pm - 2:45pm
Generating Mechanisms of Epileptic Seizures at the Level of Single Neurons Recorded in Humans in vivo
Presenter: Florian MORMANN, University of Bonn, Germany
The mechanisms of seizure generation, propagation, and termination in humans are poorly understood at the level of single neurons and micro-circuits. In this talk I will present data from microwire recordings in the human medial temporal lobe during interictal, preictal, and ictal periods. A particular emphasis will be placed to the differential contributions of putative (excitatory) principal cells and (inhibitory) interneurons during seizure generation and propagation. Furthermore, I will address the role of micro-seizures as potential precursors of clinical seizures and assess their predictive value.
Statistical challenges relate to the use of clustering algorithms for spike sorting of detected action potentials from different neurons and for the classification of putative interneurons vs. principal cells based on state space distributions of electrophysiological signature.

2:45pm - 3:10pm
Reliability of Electroencephalographic Measures
Presenter: Yvonne HÖLLER, University of Akureyri, Iceland
The replication crisis has raised awareness to sample size and statistical methods, but should we also reconsider the measures we are using in neurology? The electroencephalogram (EEG) is used in clinical practice in neurology, in neurosciences, and is investigated also by computer scientists as a new biometric tool for secure user authentication. However, the awareness of poor reliability of features derived from the EEG varies between these disciplines. I this talk I will give an overview of studies concerning reliability of measures derived from the EEG in neurological research, and relations to psychology and computer science. While most computer scientists estimate the usability of the EEG for brain computer interfaces due to low reliability to be unrealistic, the EEG is commonly used in neurological practice, however, by extracting other, presumably more reliable features. In contrast, recent research could show that reliability is rather low for measures of connectivity which are celebrated as an ongoing hype in neuroscience.

MT1: Recent Advances in Modelling Techniques and Hypothesis Testing Problems I
Blue Lecture Hall (HS 402)

3:10pm - 3:35pm
Robust Random Forest for Regression
Presenter: Aylin ALIN, Dokuz Eylul University, Turkey

Because of its ability to work with very large data sets and to provide additional features such as measure of variable importance random forest algorithm has become one of the most popular and powerful statistical learning methods. Random forest can be used for classification and regression. In this study, we focus on its use for regression where it is not robust to outliers. Our aim is to improve the robustness of the method. We propose new weighted approach to aggregate the trees, and to split the nodes in each tree. We also introduce the sufficient bootstrap and m out of n (moon) bootstrap into random forest to improve its consistency. We investigate the performance of the proposed robust random forest algorithm on artificial and real data sets.

3:35pm - 4:00pm

Berkson's Paradox, What Next? Extracting Information for the Entire Population: An Application to Alzheimer's Disease

Presenter: Polychronis ECONOMOU, University of Patras, Greece
Co-Authors: Apostolos BATSIDIS, George TZAVELAS, Panagiotis ALEXOPOULOS

Berkson's paradox, one of the most famous paradox in probability and statistics, occurs whenever in a study unrelated random variables become spuriously associated together in the observed sample. In 1946 Joseph Berkson first illustrated the paradox with a case–control study linking diabetes with cholecystitis amongst inpatients who seek care. The two diseases were found to be positive correlated even if they are independent in the population. Berkson himself explained this spuriously finding by recognizing a patient with more than one disease was more likely to be hospitalized than a patient with only a single disease.

Since then, many authors encourage physicians to understand Berkson's paradox in order to avoid misinterpreting data whenever counter-intuitive findings are observed. In the present work we try to move one step forward. More specifically, we focus on how we can extract information for the entire population based on a sample in which the Berkson's paradox is observed. This is done by recognizing firstly that Berkson's paradox is actually a selection bias problem introduced by the sampling procedure and secondly by using proper tools to describe biased samples. These tools are related to the concept of weighted distributions and to the Approximate Bayesian Computation rejection algorithm, a likelihood-free inference method. Moreover, some remarks are given regarding the selection among candidate models.

The proposed method is illustrated in a real data application for patients with dementia due to Alzheimer's disease. The application reveals characteristics of the population that are masked by the sampling procedure

4:00pm - 4:25pm

Divergence-Based Inference for Multinomial Distributions with Simulations

Presenter: Christos MESELIDIS, UNIVERSITY OF THE AEGEAN, Greece
Co-Authors: Alex KARAGRIGORIOU, Ilia VONTA

Measures of divergence or discrepancy are used extensively in statistics in various elds. In statistics, the problem of determining the appropriate distribution or the appropriate model for a given data set is extremely important for reducing the possibility of erroneous inference.
In the present work we focus on estimation and hypothesis testing and based on a general family of measures we propose a general family of test statistics that involves two indices, the values of which play a key role in the eectiveness of the proposed methodology. The asymptotic properties of the associated estimators are also examined. Finally, we explore through extensive simulations, the effect of the shape of the multinomial distribution on the performance of the proposed test.

NP1: Nonparametric Inference I
Green Lecture Hall (HS 403)

3:10pm - 3:35pm
Non-parametric Archimedean Generator Estimation With Implications For Multiple Testing
Presenter: Thorsten DICKHAUS, University of Bremen, Germany
Co-Authors: André NEUMANN
In multiple hypothesis testing, the family-wise error rate can, under some conditions, be bounded by the copula of the test statistics. Assuming that this copula is Archimedean, we consider two non-parametric Archimedean generator estimators. The first estimator is taken from the literature, while the second one is a novel modification thereof. In simulations, we compare the resulting multiple tests with the Bonferroni test and the multiple test derived from the true generator as baselines.

3:35pm - 4:00pm
Process of R-Estimators of Slopes in Linear Model
Presenter: Jana JURECKOVA, The Czech Academy of Sciences, Institute of Information Theory and Automation, Czech Republic
We consider the family of R-estimators in the linear regression model with independent errors, identically distributed according to an unknown distribution function F. The R-estimators are based on Hájek's rank scores $a(i, \alpha)$, $i = 1, \ldots, n$, The estimator $\beta(n\alpha)$ of β, defined as the minimizer of the Jaeckel rank dispersion, is after standardization with density quantile function and with the regression matrix, asymptotically equivalent to a specific rank statistic. For $0 \leq \alpha \leq 1$, this transformation converges to the Brownian Bridge. Hence, the R-estimators $\beta(n\alpha)$ have different rates for various α and can be consistent even for $\alpha \uparrow 1$.

4:00pm - 4:25pm
Edgeworth's Time Series Model: Not AR(1), but the same Covariance Structure

Presenter: Stephen PORTNOY, University of Illinois, United States of America
Autoregressive models have been popular choices for modeling time series data since the early work of Yule (1927), especially when a scatter plot of pairs of successive observations indicate a linear trend. Curiously, some 40 years earlier, Edgeworth developed a different model for treating economic time series with random increments and decrements. A version of this process with normal errors has the same covariance structure as an AR(1) process, but is actually a mixture of a very large number of processes, some of which are not stationary. That is, joint distributions of lag 3 or greater are not normal but are mixtures of normals (even though all pairs are bivariate normal). This Edgeworth Process has many additional surprising features, two of which are: (1) it has Markov structure, but is not generated by a one-step transition operator, and (2) the sample paths look very much like an AR(1), but it can be distinguished from an AR(1) about as well as distinguishing a mean difference of nearly 1 standard deviation with normal samples of size 100 or greater. It is widely recognized that model identification and verification are needed to avoid serious errors in inference. Examples like this one show that standard model-fitting diagnostics (like any ones based on second order properties) can be entirely inadequate and misleading.

FC1: Free contributions I
Green Lecture Hall (HS 403)

4:50pm - 5:10pm
Marked Linear Rank Statistics
Presenter: Ulrich MÜLLER-FUNK, Westfälische Wilhelms-Universität Münster, Germany
Co-Authors: Stefanie WEISS
The paper deals with statistics generalizing one-sample linear signed rank statistics. Formally, the sign of some quantitative variable X is now replaced by a dummy that captures the levels of a quantitative variable Y. That class naturally arises when testing the hypothesis of X/Y-independence. Idea: Compute X-ranks and check if large/small ranks go along with some Y-levels. If no such pattern is perceptible, decide in favor of the hypothesis. Assessing the dependence of variables measured on mixed scales, is a long standing problem. The present approach avoids binning and shies away from doubtful attempts to employ two-sample procedures. As for applications, feature selection in the context of classification problems by means of (multiple) testing will be mentioned. A focus of the contribution is on the basic distributional theory. For the sake of simplicity, the presentation is essentially restricted to a binary Y.

5:10pm - 5:30pm
Tests For Normality Based On Approximated Probability-Weighted Empirical Transforms
Presenter: Tomasz HOLYNSKI, Vienna University of Technology, Austria
Goodness-of-fit tests based on transforms, such as characteristic functions and Laplace transforms, are regularly developed and improved over recent years. In particular, a new

research path is paved by the notion of the probability-weighted characteristic function (PWCF) introduced by S. Meintanis and his collaborators (2014, 2016). In a classical transform-based test, the test statistic has the form of L-2 type functional that measures the discrepancy between the model- and the empirical transform. To maximize the power against specific alternatives, its integrand is equipped with a function downweighting the discrepancy in appropriate regions of the transform domain. The motivation behind the PWCF is to alleviate the problem of the optimal choice of that weight function by introducing a data-driven weight already 'within' the transform. Unfortunately, the expressions for parametric PWCFs of the standard densities are not available in closed forms. Hence, the wide-spread use of the method may be inhibited by necessity of double numerical integration: to obtain the values of the transform and those of the test statistic. In this study, we focus on testing for normality in the proposed way. To enlarge our toolkit, we introduce also the probability weighted Laplace transform (PWLT) and discuss its properties. We stress that while the standard bilateral Laplace transform of the normal density is unbounded, the corresponding PWLT is bounded; this is important for many procedures in which the always bounded characteristic function has been favored. Next, we show that in the normal case both the probability-weighted transforms can be approximated so that the test statistics are easier to compute. With this advantage, large-scale simulations are conducted for power assessments and comparisons that are missing in literature. As the distributions of the test statistics are hard to derive analytically, to estimate the critical points the parametric bootstrap is used.

5:30pm - 5:50pm

Applying Sequential Testing when Standardization of a Psychological Test needs a Checkup

Presenter: Larissa BARTOK, Modul University Vienna, Austria
Co-Authors: Klaus D. KUBINGER, Jan STEINFELD

The DIN 33430 demands a checkup of a psychological test standardization's appropriateness at least every eight years. Because such checkups require large sample sizes and are therefore very expensive there is a need to find the minimum required sample size for standardization according to a given type-I- and type-II-risk and a certain effect of model contradiction when testing the Rasch model. This contribution introduces a new algorithm using the sequential triangular two-sample t-test to test the Rasch model's quality of specific objective measurement sequentially. A simulation study emphasises the usefulness of the new approach comparing the sequential to the usual approach using an LR-test and the full sample size. Furthermore, the number of items needed is estimated as well as the number r* risking a type-I-error is empirically determined. The results of a 10-item and a 60-item setting with no DIF and two different DIF sizes are discussed in the presentation.

5:50pm - 6:10pm

How to Assess Rater Rankings? A Theoretical and Simulation Approach Using the Sum of the Absolute Pairwise Row Differences (PARDs)
Presenter: Matthias Alexander BURZLER, University of Applied Sciences Wiener Neustadt, Austria
Co-Authors: Larissa BARTOK
Although the evaluation of inter-rater agreement is often necessary in psychometric procedures (e.g. standard-settings or assessment centers), the measures typically used are not unproblematic. Cohen's Kappa and Fleiss' Kappa are known for penalizing raters in specific settings, are highly dependent on the marginals and are not a useful choice in ranking settings. This contribution introduces a new approach filling this gap by using the probability of consistencies in a setting where n independent raters rank k objects. The discrete theoretical probability distribution of the sum of the pairwise absolute row differences (PARDs) is used to evaluate rater agreement of empirically retrieved rating results. This is done by calculating the sum of PARDs in an empirically obtained nxk matrix together with the theoretically expected distribution of the sum of PARDs assuming raters randomly ranking items. The presentation covers both theoretical and applied considerations of the PARDs-approach. A simulation study highlights the usefulness of the approach and its performance by comparing existing measures to the new approach.

MT2: Recent Advances in Modelling Techniques and Hypothesis Testing Problems II
Auditorium Maximum (HS 401)

4:50pm - 5:15pm
Effect of Nonnormality on the Distribution of the Statistics for Testing a Mean Vector with Two-step Monotone Missing Data
Presenter: Nobumichi SHUTOH, Kobe University, Japan
Supposing that two-step monotone missing data is observed under elliptically contoured pattern-mixture model, we derive asymptotic distributions of the statistics for testing a mean vector: Hotelling's T^2-type test statistic and the likelihood ratio test statistic. The main results help us to investigate the effect of nonnormality on the null distribution of the test statistics. We also show the results obtained by performing Monte Carlo simulation in order to evaluate the validity of our results under some cases.

5:15pm - 5:40pm
Asymmetry Models for Square Contingency Tables with Ordinal Categories
Presenter: Kouji TAHATA, Tokyo University of Science, Japan
For the analysis of square contingency tables with ordinal categories, the issues of symmetry rather than independence arise naturally. The symmetry model indicates the structure of symmetry of cell probabilities and is expressed by the log-linear model. Also, various asymmetry models, which are expressed by the log-linear model, have been proposed. For example, the extended linear asymmetry model is proposed by Tahata, Naganawa and Tomizawa (2016). Kateri and Papaioannou (1997) and Kateri and Agresti (2007)

proposed the class of asymmetry models based on the f-divergence. Fujisawa and Tahata (2018) proposed the asymmetry models based on logit transformations.
In this talk, the relationships between these models are mentioned. When the symmetry model fits poorly for the real dataset, the decomposition of symmetry may be useful to see the reason of it. Thus, the decomposition of symmetry and the properties of test statistics are discussed.

5:40pm - 6:05pm
Testing Identity And Sphericity For Covariance Matrix For High-dimensional Data Under General Distribution
Presenter: Takayuki YAMADA, Kagoshima University, Japan
Co-Authors: Tetsuto HIMENO
The statistical inference for high-dimensional covariance matrix has studied by many researchers, extensively. For example, Ledoit and Wolf (2002, Anals of Statist.) examined two classical tests (John (1972), Sugiura (1972), Nagao (1973)) for sphericity and identity of the covariance matrix for high-dimensional case under normality. Chen et al. (2010, JASA) has proposed a tests which is based on U-statistic under general distribution. Zhou et al. (2014, Biometrika) has proposed high-dimensional sphericity test based on multivariate sign, which is applicable for elliptical population. However, their results have restrictive for population, i.e., the assumption for population is strong. The population distribution assumption should be weaker for high-dimensional data. So, in this study, we propose more robust tests. By demonstrating numerical simulations for attained significance level(ASL) and empirical power(EMP), we find the tendency that our proposed test are precise than any other two tests for ASL, and EMP is almost the same as Chen et al.'s test.

OD1: Optimal Design of Experiments
Blue Lecture Hall (HS 402)

4:50pm - 5:15pm
Efficiency Analysis for the Misspecificacion Problem of Response Probability Distribution in Optimal Experimental Design
Presenter: Victor CASERO-ALONSO, University of Castilla-La Mancha, Spain
Co-Authors: Sergio POZUELO-CAMPOS, Mariano AMO-SALAS
In the optimal experimental design, a homoscedastic normally distributed response is generally assumed. In some works, other probability distributions for the response are considered. Here, we analyze, in terms of efficiency, the influence of a misspecification of the response probability distribution. We rely on the elementary information matrix, which allows us to generalize Fisher's information matrix, to obtain approximate optimal designs for responses with any probability distribution. We have proved some theoretical results that quantify the loss of efficiency when the real distribution is the Gamma or Poisson distribution, but it is considered an equivalent heteroscedastic normal distribution. We provide two interactive applets, developed with the Mathematica software, to build D-optimal designs and to calculate the efficiency loss when there is a misspecification of the

response probability distribution. In addition, we have obtained optimal designs and have analyzed three real cases with models that explain: a) the time until a jam is produced in a silo, b) the relationship between the dose of insecticide and the death of an insect, and c) the effect of some drugs on cell growth in vitro through the 4-parameter Hill model.

5:15pm - 5:40pm

Excess And Saturated D-optimal Designs For The Rational Model

Presenter: Petr Valerievich SHPILEV, St. Petersburg State University, Russian Federation
Co-Authors: Yuri Dmitrievich GRIGORIEV, Viatcheslav Borisovich MELAS

The problem of constructing nonsingular *saturated* optimal designs (i.e. optimal designs with the number of support points which is equal to the number of parameters) is quite important since the use of such designs allows to decrease experimental expenses. On the other hand, *excess* optimal designs (i. e. optimal designs with the number of support points which is greater than the number of parameters) are useful in practice too, since they can be used to verify the adequateness of the model. For a rational two-dimensional nonlinear in parameters model used in analytical chemistry, we investigate how homothetic transformations of the design space affect the number of support points in the optimal designs. We show that there exist two types of optimal designs: a *saturated* design and an *excess* design. The saturated optimal designs are constructed explicitly. Numerical methods for constructing excess optimal designs are used.

5:40pm - 6:05pm

Locally Optimal Designs For Generalized Linear Models With Application To Gamma Models

Presenter: Osama IDAIS, Otto-von-Guericke- Universität Magdeburg, Germany
Co-Authors: Rainer SCHWABE

The generalized linear model is considered as a unified class of regression models for univariate continuous and discrete responses. For instance, Poisson, logistic and gamma models can be considered. In this work we derive locally optimal designs for a wide class of generalized linear models with binary factors. The results cover many work in the literature. In particular, the generalized linear models for gamma-distributed outcomes are adopted. These so-called gamma models are employed for outcomes that are non-negative, continuous, skewed and heteroscedastic, specifically, when the variances are proportional to the square of the means. Many real life data from psychology, ecology or medicine can be fitted by gamma models. Despite of that, much attention has not been given to gamma models in optimal designs considerations. In the sparse literature on this topic mostly geometric approaches were used to identify locally optimal designs.
In the current work, analytic solutions for locally optimal designs on a compact experimental region are developed under gamma models. The experimental region can be reduced to its vertices under particular model assumptions and thus, the support of a design is chosen among the vertices. The robustness of the derived designs with respect

to misspecifications of the initial parameter values is examined by means of their local efficiencies.

Wednesday September 4, 2019

EDA1: Experimental Design and Application
Green Lecture Hall (HS 403)

8:30am - 8:55am

Design Of Experiments And Machine Learning Methods To Improve Robustness Of Big Data Analytics With Application To A Real Case Study

Presenter: Luca PEGORARO, Università degli Studi di Padova, Italy
Co-Authors: Rosa ARBORETTI, Riccardo CECCATO, Laura MARTINELLO, Luigi SALMASO

When deploying predictive analytics in a Big Data context, some concerns may arise regarding the validity of the results obtained. The reason for this is linked to flaws which are intrinsic to the nature of the Big Data Analytics methods themselves. For this reason a new approach is proposed with the aim of mitigating new problems which arise. This novel method consists of a two-step workflow in which a Design of Experiments (DOE) study is conducted prior to the usual Big Data Analytics and machine learning modeling phase. The advantages of the new approach are presented and a real industrial application is introduced also to emphasize pros and cons of most popular machine learning techniques and steps needed to implement them.

8:55am - 9:20am

Testing Interaction in Different Kinds of Block Designs Based on a Linear Model

Presenter: Karl MODER, Universty of Natural Resources and Life Sciences, Austria

Block designs are often used designs to evaluate influences of a factor in the presence of some disturbance variables. Although this kind of design is widely used, it suffers from one drawback. As there is only one observation for each combination of a block and factor level it is not possible to test interaction effects, because the mean square value for interaction has to serve for the error term.
Although there are some attempts to overcome this problem these methods however, have not been adopted in practice and have not been broadly disseminated. Many of these tests are based on nonlinear interaction effects (e.g. Tukey 1949, Mandel 1961, ...). Others are based on the sample variance for each row in the block design (Milken an Ramuson 1977). Here a new model is introduced to test interaction effects in block designs. It is based on one additional assumption regarding the columns of the block design which is intuitive and common in Latin Squares. The application of this model is very simple and a test on interaction effect is very easy to calculate based on the results of an appropriate analysis of variance. The method as such is applicable for fixed effect models as well as for a certain class of mixed and random effect models.

Mandel, J. (1961). Non-additivity in two-way analysis of variance.
Journal of the American Statistical Association 56, 878- 888.
Millken, G. A. and D. Rasmuson (1977). A heuristic technique for testing for the presence of interaction in nonreplicated factorial experiments. Australian Journal of Statistics 19 (1), 3238.
Tukey, J. W. (1949). One degree of freedom for non-additivity. Biometrics 5, 232-242.

9:20am - 9:45am

Group Sequential Designs Revisited

Presenter: Sergey TARIMA, Medical College of Wisconsin, Wauwatosa, Wisconsin, United States of America
Co-Authors: Nancy FLOURNOY

Group sequential procedures are designed with a possibility of early stopping for futility or efficacy. The distributions of common fixed sample size maximum likelihood estimates (fMLE) change when possibility of early stopping is introduced. For the majority of popular groups sequential designs (including Pocock, O'Brien & Fleming, Haybittle-Peto), the distributions of these fMLE are shown to be mixtures of truncated distributions. Truncation is caused by the conditioning criteria (the requirement to stay within a continuation region) which must be met for the experiment to continue. We compare the conditional MLEs with fMLEs using a two stage Pocock design. We propose a group sequential design which controls predetermined marginal power at each of several alternatives hypotheses while controlling the type I error. The number of alternative hypotheses defines the maximum number of stages. In this new design, stage specific sample sizes are calculated to reach predetermined power for each of the alternatives.

9:45am - 10:10am

Optimum Design Of Experiments Based On Precision Requirements

Presenter: Bernhard SPANGL, University of Natural Resources and Life Sciences, Vienna, Austria
Co-Authors: Lenka FILOVA, Radoslav HARMAN

Let us assume that the observed response of each trial of an experiment depends on a design point x chosen from a design space $\mathfrak{X}$. Let Ξ_n^E denote the set of all exact experimental designs of size n on $\mathfrak{X}$. We first aim at finding the minimum experimental size n^* subject to some precision requirements δ:

$$n^* \in \arg\min n : \exists \boldsymbol{\xi} \in \Xi_n^E \text{ with } \Psi(\boldsymbol{xi}) \leq \delta,$$

where $\Psi(\boldsymbol{\xi})$ is a precision function, such that smaller values of Ψ indicate greater precision. Then we compute the optimum precision design $\boldsymbol{\xi}^*$ as

$$\boldsymbol{\xi}^* \in \arg\min \Psi(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi_{n^*}^E \text{ with } \Psi(\boldsymbol{\xi}) \leq \delta.$$

We will discuss natural choices of the precision function Psi for different types of statistical inference, i.e., estimation and testing. Next, we will suggest algorithms for constructing

optimal or nearly-optimal precision designs. Finally, we will give some examples and back our proposal by Monte Carlo simulations.

FS1: Forensic Statistics
Blue Lecture Hall (HS 402)

8:30am - 8:55am

Case Study Validations of Automatic Bullet Matching

Presenter: Heike HOFMANN, CSAFE, Ames IA, United States of America
Co-Authors: Susan VANDERPLAS

Recent advances in microscopy have made it possible to collect 3D topographic data, enabling virtual comparisons based on the collected 3D data next to traditional comparison microscopy. Automatic matching algorithms have been introduced for various scenarios, such as matching cartridge cases (Tai and Eddy 2018) or matching bullet striae (Hare et al. 2017, Chu et al 2013, De Kinder and Bonfanti 1999). One key aspect of validating automatic matching algorithms is to evaluate the performance of the algorithm on external tests. Here, we are presenting a discussion of the performance of the matching algorithm (Hare et al. 2017) in three studies. We are considering matching performance based on the Random forest score, cross correlation, and consecutive matching striae (CMS) at the land-to-land level and, using Sequential Average Maxima scores, also at the bullet-to bullet level. Cross correlation and Random Forest scores both result in perfect discrimination of same-source and different-source bullets. At the land-to-land level, discrimination (based on area under the curve, AUC) is excellent (≥ 0.90).

8:55am - 9:20am

Bayesian Characterizations Of U-processes Used In Pattern Recognition With Application To Forensic Source Identification

Presenter: Cami Marie FUGLSBY, South Dakota State University, United States of America

In forensic science, a typical interpretation task is a common-but-unknown-source identification, where an examiner must summarize and present the evidential value associated with two sets of objects relative to two propositions. The first proposition is that the two sets of objects are two simple random samples from the same, unknown source in a given population of sources; the second proposition is that the two sets of objects are two simple random samples each drawn from two different but unknown sources in a given population of sources. Typically, the examiner has to develop criteria or a rule to compare the two sets of objects; this rule leads to a natural U-process of degree two for assessing the evidence. In this work, we will characterize the U-process and demonstrate how to write a class of approximately admissible decision rules in terms of the U-process. Combining the asymptotic representation of this U-process with an approximate ABC algorithm, we can then provide summary statistics with Bayes factor-like properties for the selection between the two propositions. We will illustrate this method with an application based on recovered aluminum powders associated with IEDs.

For complex evidence forms, we usually have to learn the metric for comparing two samples. Typically, there is not a natural feature space for which modern statistical techniques can be applied to the non-nested models for model selection. In this presentation, a score function has been developed that maps the trace samples from their measured feature space to the real number line. The resulting score for two trace samples can be used as a measure of the atypicality of matching samples, which will be applied to a receiver operating characteristic (ROC) curve and in a score-based likelihood ratio function.

9:20am - 9:45am

Which Forensic Likelihood Ratio Approach is Better?: An Information-Theoretic Comparison

Presenter: Danica OMMEN, Iowa State University, United States of America
Co-Authors: Peter VERGEER

There are several methods for constructing likelihood ratios (LR) for forensic evidence interpretation. Feature-based LR approaches directly model the measured features of the evidential objects while score-based LR approaches model the similarity (or sometimes the dissimilarity) between two objects instead. The score-based approaches often rely on machine learning methods of producing the similarity scores. In addition to how the evidence is treated, the LR approaches also differ in the type of propositions (or hypotheses) they address. In this presentation, we will only consider source-level propositions that address the origin of a particular set of evidence, regardless of the actions or motivations involved. In particular, we consider the common-source and the specific-source propositions. It has been shown that the different propositions and treatments of the evidence lead to differing values of the computed LR. So, which method is preferred for the interpretation of forensic evidence? We will use methods from information theory to compare the various LR approaches.

9:45am - 10:10am

ROC Curves And Frequentist/Machine-Learning Based Likelihood Ratios For Source Identification

Presenter: Larry TANG, george mason university, United States of America
Co-Authors: Danica OMMEN, Elham TABASSI, Xiaochen ZHU

The likelihood ratio based on similarity scores recently brings attention to the forensic scientists, especially on the automated facial recognition system scores on faces. National Institute of Standards and Technology publishes comprehensive reports on the performance of the commercial matching algorithms. As the algorithms for matching facial images are largely proprietary, it is easier to obtain the similarity scores than the original configurations used in the algorithms. The purpose of this talk is to introduce the score-based likelihood ratio based on receiver operating characteristic (ROC) curve analysis. The ROC curve is widely used in radiology, psychophysical and medical imaging research for detection performance, military monitoring, and industrial quality control. We will introduce methods for estimating the likelihood ratio from the ROC curve that

is estimated with machine learning techniques for source identification, and derive the confidence interval for the likelihood ratio.

10:10am - 10:35am
Discussion Of Presentations In The Forensic Science Session
Presenter: Sonja MENGES, Bundeskriminalamt, Germany
Co-Authors: Alicia CARRIQUIRY
Sonja Menges and Alicia Carriquiry are serving as discussants at the forensic science session.

FD1: Functional Data
Green Lecture Hall (HS 403)

11:00am - 11:25am
Nonparametric Density Estimation For Intentionally Corrupted Functional Data
Presenter: Alexander MEISTER, Universität Rostock, Germany
We consider statistical models where functional data are artificially contaminated by independent Wiener processes in order to satisfy privacy constraints. We show that the corrupted observations have a Wiener density which determines the distribution of the original functional random variables uniquely, and we construct a nonparametric estimator of that density. We derive an upper bound for its mean integrated squared error which has a polynomial convergence rate, and we establish an asymptotic lower bound on the minimax convergence rates which is close to the rate attained by our estimator. We discuss data-driven ways of choosing the basis and the smoothing parameters and provide numerical simulations. This talk is based on a joint work with Aurore Delaigle (University of Melbourne, Australia).

11:25am - 11:50am
A Functional Version Of The ARCH(p) And GARCH(p,q) Model
Presenter: Sebastian KÜHNERT, University of Rostock, Germany
The ARCH(p) model and the GARCH(p,q) model with integers p,q are usually applied in order to model real valued financial time series and will be established for time series which take their values in function spaces of measurable functions on the domain [0,1]. Sufficient conditions for the existence of strictly stationary solutions will be provided for $L^p[0,1]$ spaces and separable function spaces with domain $[0,1]$ endowed with the supremum norm. Estimators will be constructed for the shift term and for the coefficient operators of the functional ARCH(p) equation and of the functional GARCH(p,q) equation as well. For these estimators asymptotic upper bounds with an explicit convergence rate are deduced in the sense of convergence in probability.

11:50am - 12:15pm

Mode Estimation For Functional Data
Presenter: Dennis MÜLLER, Universität Rostock, Germany
In the case of a finite dimensional probability distribution that has a density function with respect to the Lebesgue measure, the mode is often defined as a value that maximizes the probability density function. Unfortunately, in the case of i.i.d. functional random variables one does not have a Lebesgue-density function. In this lecture we are going to extend the notion of the mode of a probability distribution to distributions on infinite dimensional spaces, i. e. function spaces. Therefore we shall firstly assume the existence and uniqueness of the mode. We will consider examples of some functional distributions (i.e. Gaussian measures) and discuss the availability of a (unique) mode. Relying on entropy inequalities and coverings, an estimator of the mode is constructed. We then prove its consistency and discuss rates of convergence.

12:15pm - 12:40pm
Tests and Confidence Regions for Incompletely Observed Functional Data
Presenter: David KRAUS, Masaryk University, Brno, Czech Republic
We study methods for the analysis of functional data under partial observation, by which we mean situations, where each functional variable may be observed only on a subset of the domain while no information about the function is available on the complement. Interestingly, some essential methods, such as K-sample tests of equal means or covariances and confidence intervals for eigenvalues and eigenfunctions, that are well established for completely observed curves, are lacking under the incomplete observation regime. The only currently available approach, in which incomplete curves are omitted, is clearly suboptimal and even infeasible, if there are no complete curves. We study methods that use all curve fragments and do not even require any complete curves. The principal difficulty in the practical implementation is the impossibility to perform dimension reduction, resulting in large objects that are often impossible to store in computer memory and perform computation with. The bootstrap turns out to be a way to address this problem. Theory, simulations and a data example will be presented.

NP2: Nonparametric Inference II
Blue Lecture Hall (HS 402)

11:00am - 11:25am
Goodness Of Fit And Stochastic Ordering Permutation Tests
Presenter: Luigi SALMASO, Università degli Studi di Padova, Italy
Co-Authors: Rosa ARBORETTI GIANCRISTOFARO, Riccardo CECCATO, Laura MARTINELLO, Luca PEGORARO
This work focusses the attention on non parametric combinations (NPC) of dependent permutation tests for ordered categorical variables. In particular, this methodology is applied on goodness of t tests and stochastic dominance tests for restricted alternatives. Several parametric solutions to univariate case have been proposed in literature. However, these methods, most of which are based on the restricted maximum likelihood ratio test,

are generally criticized because of their dependence on the unknown population distribution F and the difficulty or impossibility to use them in multivariate cases. Some of them, due to the quadratic form, allow two-tailed tests to verify dissimilarity, but they cannot permit hypothesis testing on stochastic dominance. NPC solution can be applied on every type of variable and it is valid also in case the sample sizes are lower than the number of variables.

11:25am - 11:50am

Testing for Equivalence by Combined Permutation Tests

Presenter: Fortunato PESARIN, University of Padova, Italy
Co-Authors: Luigi SALMASO, Rosa ARBORETTI

The notion of testing for equivalence of two treatments is widely used in clinical trials, pharmaceutical experiments, bioequivalence and quality control. It is traditionally operated within the intersection-union principle (IU). According to this principle the null hypothesis is stated as the set of effects the differences δ of which lie outside a suitable equivalence interval and the alternative as the set of δ that lie inside it. In the literature related solutions are essentially based on likelihood techniques, which in turn are rather difficult to deal with. Recently, an approach based on Roy's union-intersection principle (UI) has been published that, staying within the permutation theory, goes beyond some limitations of likelihood based methods. The UI approach, effectively a mirror image of IU, assumes a null hypothesis where δ lies inside the equivalence interval and an alternative where it lies outside. Since testing for equivalence can rationally be analyzed by both principles, but since the two differ in terms of the mirror-like roles assigned to the hypotheses under study they are not strictly comparable. The present paper's main goal is to looking into these problems by providing a sort of parallel analysis of both and by highlighting the related requirements, properties, limitations, difficulties, and pitfalls so as to get practitioners properly acquainted with their use in practical contexts.

11:50am - 12:15pm

The Limiting Distribution of Combining the t and Wilcoxon Rank Sum Tests

Presenter: Masato KITANI, Tokyo University of Science, Japan
Co-Authors: Hidetoshi MURAKAMI

The two-sample testing problem is one of the most important topics in testing statistical hypothesis. Student's t test and Wilcoxon's rank sum test are often used to test the location parameter, and these tests have been discussed by many authors over the years. Combining these two tests within a maximum test has been considered. It was shown that the maximum test controls the type I error and has good power property for various distributions. However, the limiting distribution of the maximum test has not been derived prior to this publication. The limiting distribution of the maximum test is derived under the null hypothesis. Simulations are used to investigate the convergence of the maximum test to the limiting distribution for various cases.

12:15pm - 12:40pm

The Generalized Multisample Cucconi Statistic for the Location and Scale Parameters

Presenter: Takuya NISHINO, Tokyo University of Science, Japan
Co-Authors: Hidetoshi MURAKAMI

Many researchers suggested various test statistics for dealing with data appropriately. A nonparametric one-way layout analysis of variance plays an important role in biometry. Various test statistics have been proposed to address the multisample location-scale problem. We propose a generalized multisample Cucconi test statistic for the location, scale and location-scale parameters. The null and non-null limiting distribution of the generalized multisample Cucconi test is derived. Deriving the exact critical value of the test statistic can be difficult when the sample sizes are increasing. An approximation method to the distribution function of the test statistic can be useful in this situation. We use a gamma approximation to evaluate the upper-tail probability of the suggested test statistic for the finite sample size. Simulation studies are used to investigate the convergence of the generalized multisample Cucconi test statistic to the limiting distribution under various scenarios.

SL1: Statistical Learning - Methods and Applications I
Auditorium Maximum (HS 401)

11:00am - 11:40am

Scaled Expected Improvement for Bayesian Optimization

Presenter: Umberto NOÈ, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
Co-Authors: Dirk HUSMEIER

Bayesian optimization (BO) is a popular algorithm for solving challenging optimization tasks. It is designed for problems where the objective function is expensive to evaluate, perhaps not available in exact form, without gradient information and possibly returning noisy values. Different versions of the algorithm vary in the choice of the acquisition function, which recommends the point to query the objective at next. Initially, researchers focused on improvement-based acquisitions, while recently the attention has shifted to more computationally expensive information-theoretical measures. In this talk I will present two major contributions to the literature. First, I propose a new improvement-based acquisition function that recommends query points where the improvement is expected to be high with high confidence. The proposed algorithm is evaluated on a large set of benchmark functions from the global optimization literature, where it turns out to perform at least as well as current state-of-the-art acquisition functions, and often better. This suggests that it is a powerful default choice for BO. The novel policy is then compared to widely used global optimization solvers in order to confirm that BO methods reduce the computational costs of the optimization by keeping the number of function evaluations small. The second main contribution represents an application to precision medicine, where the interest lies in the estimation of parameters of a partial differential equations model of the human pulmonary blood circulation system. Once inferred, these

parameters can help clinicians in diagnosing a patient with pulmonary hypertension without going through the standard invasive procedure of right heart catheterization, which can lead to side effects and complications (e.g. severe pain, internal bleeding, thrombosis).

11:40am - 12:00pm

Correlated Parameters to Accurately Measure Uncertainty in Deep Neural Networks

Presenter: Konstantin POSCH, University Klagenfurt, Austria
Co-Authors: Jürgen PILZ

In this article a novel approach for training deep neural networks using Bayesian techniques is presented. The Bayesian methodology allows for an easy evaluation of model uncertainty and additionally is robust to overfitting. These are commonly the two main problems classical, i.e. non-Bayesian, architectures have to struggle with. The proposed approach applies variational inference in order to approximate the intractable posterior distribution. In particular, the variational distribution is defined as product of multiple multivariate normal distributions with tridiagonal covariance matrices. Each single normal distribution belongs either to the weights, or to the biases corresponding to one network layer. The layer-wise a posteriori variances are defined based on the corresponding expectation values and further the correlations are assumed to be identical. Therefore, only a few additional parameters need to be optimized compared to non-Bayesian settings. The novel approach is successfully evaluated on basis of the popular benchmark datasets MNIST and CIFAR-10.

12:00pm - 12:20pm

Application Of Supervised Learning To Predict The Quality Of A Welding Process

Presenter: Kathrin PLANKENSTEINER, FH Vorarlberg, Austria
Co-Authors: Melanie ZUMTOBEL

In reliability analysis, End-of-Life tests are necessary to guarantee that products operate reliably. Since it is not possible to test all the devices at real stress conditions, accelerated stress tests in combination with statistical models are commonly applied to achieve reliable forecasts for the lifetime.

In this paper, the statistical framework of lifetime modeling and prediction is discussed with application to an industrial setting, where the welding procedure is known to be the critical step during production. As an example case, lifetime data from a device made of sheet metal has been collected. To test the quality of the welded area, End-of-Life tests have been performed on each of the devices.

For lifetime modeling, commonly applied lifetime distributions, e.g., Weibull and log-normal, are discussed, evaluated and compared using expert knowledge, but also statistical measures of goodness-of-fit like the ks-test and the Bayesian Factor.

To model the lifetime depending on the welding parameters, a multivariable linear regression model is applied. To find the significant covariates, a mix of forward selection and

backward elimination is used. For model evaluation, cross-validation and sum of squared errors are used to evaluate the predictive power of the model.
The results show that the lifetime can be modeled based on welding settings. For lifetime prediction, the model yields accurate results when interpolation is used. However, an extrapolation beyond the range of available data shows the limits of a purely data-driven model.

12:20pm - 12:40pm
On Machine Learning In Regression Analysis
Presenter: Svetlana Nikolaevna LEORA, St. Petersburg State University, Russian Federation
Co-Authors: Sergey Michaylovich ERMAKOV
As is known, the task of constructing a regression function from observed data is of great practical importance. In the case of additive error of observations at points whose coordinates are given without errors, we have: $y_j = f(X_j) + \varepsilon_j$, where $j = 1, \ldots, N$ is the observation number, y_j – the observed value, $X_j = (x_{1j}, \ldots, x_{sj})$ is the point at which the observation took place, ε_j is the observation error. It is also assumed $E\varepsilon_j = 0$. The task is to define a function f that is usually considered to be given parametrically, $f(X) = f(X, U)$, where U are unknown parameters. The problem has obvious connections with problems of approximation of functions.
The report discusses one of the possible approaches using the idea of machine learning. It is based on the approximation problem for some function f. Let A be a linear operator acting in a linear normed space F. If A^* is an adjoint operator to A, the functions ϕ_j and ψ_j of F satisfy the conditions: $AA^*\phi_j = s_j^2\phi_j$, $AA^*\phi_j = s_j^2\psi_j$, $j = 1, \ldots, r(A)$, then among all m-dimensional ($m \leq r(A)$) operators A_m the operator $\tilde{A}_m = \sum_{j=1,m} s_j(\cdot, \phi_j)\psi_j$ minimizes the norm of $||A - A_m||$.
If $K = I - A$ is an operator, such that $Kf = 0$, and in this equality we replace A with its approximation $\tilde{A}_m$, then we get an approximation to f in the form $\hat{f} = \sum_{j=1,m} s_j(f, \phi_j)\psi_j$.
The idea of ""learning" is as follows. Let$K(\theta)$ – the parametric family of operators. Using sampled values of f, we find an operator $K_0 = K(\theta_0)$, $\theta_0 = \arg\min ||K(\theta)||$, were θ belongs to θ. Assuming $A = I - K_0$, we find the corresponding functions ϕ_j and ψ_j and construct an approximation $\hat{f}$ for the appropriate m.
Real algorithms use spaces of functions defined at discrete points. A well-known particular case of using this approach is SVD time series analysis. Based on this approach, there are different generalizations of this analysis. Some examples of generalization are given in the report.

SM1: Statistical Modelling and Data Analysis
Green Lecture Hall (HS 403)

1:30pm - 1:55pm
Construction Of Two-phase Designs For Experiments With A Single Block Factor In Each Phase
Presenter: Hans-Peter PIEPHO, University of Hohenheim, Germany

Co-Authors: Nha VO-THANH
Design of experiments is widely used in many areas. Often, experiments involve two different phases. For example, in plant breeding the first phase of the experiment is performed in a field involving a number of treatments (i.e. varieties) and a single blocking factor (i.e. field blocks), whereas the second phase is performed in a laboratory to measure the response using the samples from the first phase, taking into account the presence of another blocking factor (i.e. days or lab machines). Such experiments are referred as to two-phase experiments. Currently, construction approaches for such experiments are mostly focused on the case when treatments are orthogonal to blocks and design construction using A-optimality can be conveniently based on approaches like the 'design key' proposed by Bailey (2016). However, there is as yet no general strategy to design generation when treatments are not orthogonal to blocks, which is often the case. In this article, we, therefore, propose three different approaches to generate such designs, which utilize a metaheuristics search method, namely, iterated Tabu search. The first one is a sequential approach in which we first obtain an optimal arrangement of treatments with respect to the first blocking factor (i.e. blocks in Phase 1). Given the resulting first-phase design, we find an optimal arrangement of this design with respect to the second blocking factor (i.e. blocks in Phase 2). The second approach is a simultaneous approach, which allows to find an optimal arrangement of treatments in the first-phase and second-phase block factors simultaneously. The third approach arranges treatments in a nested row-column structure, where rows and columns correspond to the two blocking factors. To avoid a computational burden for all approaches, we use an efficient update formula for computing the A-optimality criterion. We demonstrate the capacity of our algorithms with various examples, along with a discussion.

1:55pm - 2:20pm
Efficiently Searching High-Dimensional Polynomial Space: Algorithms And Guarantees
Presenter: Kory Douglas JOHNSON, University of Vienna, Austria
There has recently been a renewed interest in polynomial regression. Authors have considered SURE screening algorithms in the space of polynomials (Kong et al. 2017) and others have used polynomial regression as a tool to understand deep networks (Cheng et al. 2019). There are many well-known problems with polynomial regression: prediction can be poor near the boundaries, extrapolation is dubious at best, high-degree polynomials can be highly correlated, and even in rather small data sets the number of polynomials is unwieldy. This paper effectively solves the latter two problems and provides a heuristic for improving the former two. We present a sequential algorithm for searching the feature-space of polynomials which is based on the principle of marginality. Furthermore, the procedure controls a notion of false rejections by implementing Revisiting Alpha-Investing. This provides a valid stopping criterion to prevent over-selection.

2:20pm - 2:45pm

Bayesian Spectral Density Estimation For Multiple Time Series
Presenter: Annalisa CADONNA, WU Vienna University of Economics and Business, Vienna, Austria
Co-Authors: Raquel PRADO, Athanasios KOTTAS
The spectral density function contains information about the second order behavior of a signal and can be interpreted as the distribution of the power over the frequency range. The problem of estimating the spectral density function arises naturally in fields where informationabout frequency behavior is relevant and several signals are recorded concurrently. For example, multichannel electroencephalography (EEG) records measurements of electrical potential fluctuations at multiple locations on the scalp of a subject. I will present a hierarchical Bayesian modeling approach to spectral density estimation for multiple time series, where the log- periodogram of each series is modeled as a mixture of Gaussian distributions with frequency-dependent weights and mean functions. The implied model for each log-spectral density is a mixture of mean functions with frequency-dependent weights. In addition to accommodating flexible spectral density shapes, a practically important feature of the proposed formulation is that it allows for ready posterior simulation through a Gibbs sampler with closed form full conditional distributions for all model parameters. I will show results for multichannel electroencephalographic recordings, which provide the key motivating application for the proposed methodology. I will then present some extensions of the model to non-stationary time series.

2:45pm - 3:10pm
Control of the Population-Wise Error Rate in Clinical Trials with Multiple Populations
Presenter: Werner BRANNATH, University Bremen, Germany
Co-Authors: Kornelius ROHMEYER, Charlie HILLNER
In confirmatory clinical trials in which several populations are investigated simultaneously, control of the multiple type I error rate is usually considered necessary. However, if a treatment or a treatment strategy is tested in several disjoint populations, each population is effected by only a single hypothesis test. Hence, multiple testing adjustments appear unnecessarily conservative in this case. This observation leads us to define a new and more liberal concept of multiple type error control that accounts for the actual risk of the individual populations by any of the trials false rejections. The new concept should lead to an almost unadjusted testing when the population overlaps are small. To this end we suggest to control the risk for a randomly chosen, future patient to belong to a population that is exposed to an inefficient treatment by the trials final test decisions. We call this the “population-wise” error rate. We will introduce single stage and sequential two-stage multiple test procedures with control of the “population-wise” error rate and will investigate their utility in comparison to more traditional approaches by examples and simulations.

3:10pm - 3:35pm

Variance Component Estimators OPE, NOPE and AOPE In Linear Mixed Effects Models

Presenter: Subir GHOSH, University of California, Riverside, USA, United States of America

Optimum variance component estimation methods that yield uniformly minimum variance quadratic unbiased estimators for a full dataset are often difficult or impossible to implement. In this paper we propose an estimator which is near optimal under some distributional assumptions that can be made without specifying an exact functional form. This estimator has an exact closed form expression. An average optimum estimator, which can be used when optimum estimators exist for subsets of the data, is also proposed. Performance comparisons of the proposed estimators are made individually with other popular estimators using simulated data. A performance comparison of the average optimum estimator is made under four constraints on the variance components. A real dataset is analysed using the proposed estimators. The robustness properties of the proposed estimators, in comparison with the other estimators, including the method of moments, are also investigated, using data simulated from a skew normal distribution. The average optimum estimator is strongly robust and far superior for estimating one of the variance components, as demonstrated by making comparisons with other methods. These comparisons are based on bias, mean squared error, and mean absolute deviation. The average estimator is moderately robust in respect of the estimation of the other variance components.

SU1: Recent Developments in Survival Analysis
Blue Lecture Hall (HS 402)

1:30pm - 1:55pm

Multiple Weighted Logrank Tests With Broader Power Functions

Presenter: Marc DITZHAUS, TU Dortmund, Germany

Co-Authors: Sarah FRIEDRICH, Markus PAULY

Weighted logrank tests are a common tool for analyzing two-sample survival data. Since these tests are designed for specific alternatives the question arises how to choose a proper weight in practical application? We address this question by a flexible combination idea leading to testing procedures with broader power. Beside the classical two-sample testing problem we discuss testing for the superiority of one group. Right censoring within the data is allowed and may differ between the groups. Our tests are asymptotically exact under the null, consistent for fixed alternatives and admissible for a larger set of local alternatives. All these theoretical properties can be transferred to a permutation and wild bootstrap version, respectively. The tests' applicability is illustrated in a real data example by using our novel R-package mdir.logrank.

1:55pm - 2:20pm

A Proportional Hazards Model Under Bivariate Censoring And Truncation

Presenter: Marialuisa RESTAINO, University of Salerno, Italy

Co-Authors: Hongsheng DAI, Chao HUANG, Miriam J. JOHNSON

The bivariate survival data are usually subject to incomplete information due to censoring and truncation. Most existing works focused on estimating the bivariate survival function when only one component is censored or truncated and the other is fully observed. Only recently bivariate survival function estimation under the assumption that both components are censored and truncated has received considerable attention. Moreover, the most common approaches to model covariates effect on survival time are the Cox PH and AFT models, that have been well studied for the univariate censored data. Not much has been done for the bivariate survival data when truncation is present. The paper aims at estimating the regression coefficients in the bivariate proportional hazards model, when both components are censored and truncated. In particular, truncation is considered as covariate in the regression model, in order to evaluate its effect on the hazard estimation. A simulation study and an application on real data are conducted to investigate the performance of the estimators.

2:20pm - 2:45pm

Joint Models Of Survival And Multivariate Longitudinal Data: Diagnostic And Predictions

Presenter: Mariangela ZENGA, University of Milano-Bicocca, Italy
Co-Authors: Marcella MAZZOLENI

The joint models analyse the effect of longitudinal covariates onto the risk of an event. For studying the time-to-event a proportional hazard model is used, while the longitudinal sub-model is a linear multivariate mixed model. In this work the Expectation-Maximisation algorithm which maximises the joint likelihood function is implemented, using a one-step Newton-Raphson update and a Guass-Hermite quadrature rule. For testing the goodness of fit some diagnostics elements will be presented, such as the estimated survival function and the residuals for both, survival and longitudinal, sub-models. In addition, the dynamic predictions are introduced, for updating the survival function and the longitudinal trajectories at later points in time. The results on the well-known Primary Biliary Cirrhosis data will be shown.

2:45pm - 3:10pm

Recent Developments to Study Longitudinal and Survival Data: Applications to Biomedicine

Presenter: Ipek GULER, IMIBIC, Spain

In many observational studies, patients are followed up during a time period and several measurements are recorded in each visit. Clinicians are often interested on a time-to-event with a possible combination of biomarkers. Those biomarkers could be repeated measurements (longitudinal biomarkers) over time. To study the association between a longitudinal biomarker and a time-to-event data, there exist several techniques in the literature. Joint modeling approaches have gained an increasing attention due to their efficiency and reduction in bias. On the other hand, there are already several extensions such as to a multivariate longitudinal and time-to-event setting, multivariate survival and

longitudinal setting, among others. We aim to give an overview to joint modeling of longitudinal and survival data with biomedical data applications and software implementations.

3:10pm - 3:35pm

Exact Confidence Bounds For A Series Reliability System

Presenter: Patrick PLUM, KAI Kompetenzzentrum für Automobil- und Industrieelektronik GmbH
Co-Authors: Horst LEWITSCHNIG, Jürgen PILZ

We discuss a model that provides one-sided lower confidence intervals of the hazard rate of a series reliability system, which consists of independent components each having constant individual hazard rates. The model is applicable for a setting where time-to-failure data of components is available.
The core of the model is the combination of 'confidence distributions' of the component's hazard rates represented as rate parameters λ_i of different exponential distributions. These confidence distributions of the parameters are weighted chi-squared distributions (i.e. gamma distributions), each single providing an 'exact' upper bound for the hazard rate of the corresponding component when applying its quantiles. We are interested if the distribution obtained by the sum of these weighted chi-squared distributions, representing the series system's hazard rate $\sum \lambda_i$, retains the property of providing exact bounds. By simulation, we show that the proposed confidence procedure ensures coverage probability of at least the nominal confidence level for a variety of settings.

ENS1: Environmental Statistics
Green Lecture Hall (HS 403)

4:00pm - 4:25pm

Pollutant Dispersion Simulation By Means Of A Stochastic Particle Model And A Dynamic Gaussian-Plume Model

Presenter: Maximilian ARBEITER, Universität Klagenfurt, Austria
Co-Authors: Gunter SPÖCK, Albrecht GEBHARDT

The pollutant dispersion models of this work fall into two classes: Physical and statistical. We propose a large scale physical particle dispersion model and a dynamic version of the well-known Gaussian plume model, based on statistical lters. Both models are based on wind- measurements, -interpolations and mass-corrections of certain wind- stations installed in an alpine valley in Carinthia/Austria. Every 20 minutes the wind eld is updated and the dispersion of the pollutant is calculated. Vegetation like forest and grassland are fully taken into account. The dispersion models are used to predict pollutant concentrations resulting from the emissions of a cement plant. Both models are compared to each other and give almost equivalent results. The great advantage of the statistical model is that it does not scale like the particle model with the number of emitters, but its computational burden is constant, no matter how many emitters are included in the model. To test and validate our models, we developed the R-package PDC using the

CUDA framework for GPU implementation.

4:25pm - 4:50pm

Implementing a Low Cost Meteorological Data Network

Presenter: Albrecht GEBHARDT, Universität Klagenfurt, Austria
Co-Authors: Maximilian ARBEITER, Gunter SPÖCK

Simulation studies of environmental data usually need some meteorological data as input parameter. This is mostly available only on a rather dense grid of stations which may or may not cover the area of interest. Furthermore the available data is often only presented in some aggregated form.

Arising out of such a situation we decided to implement our own network of meteorological stations. Of course such a project had to be implemented under rather strict budget restrictions. Fortunately nowadays the availability of cheap electronic equipment like sensors or single board computers makes this possible.

We took a WH1080/WH3080 weather station as a basis. It consists of a solar-powered outdoor unit which measures wind speed and direction, temperature, humidity and for the WH3080 variant also luminosity and UV index. This outside part is connected via 433 MHz or 868 MHz radio to a battery powered indoor display unit which delivers pressure measurements. Live data can be read out via USB connection by a Raspberry PI Single board computer, connected to some WAN or equipped with its own LTE modem. Data is collected centrally in a SQL database.

In this talk we will present an overview about the client and server side implementation of reading, collecting and presenting the data. Additionally the data has to be filtered for outliers as the stations sometimes fail to communicate correctly or some sensor values are out of range. Another task is to evaluate the accuracy of these consumer grade wheather stations with professional grade hardware. For this reason we already have placed a station in cooperation with ZAMG beside their official station at Klagenfurt Airport for comparison.

Finally we show how the collected wind data can be interpolated and visualized. It is based on inverse distance weighted interpolation with some extensions dealing with mass corrections as well as terrain influence and implemented as CUDA based parallel algorithm.

4:50pm - 5:15pm

Trans-Gaussian Spatial Prediction of HCB- and Hg-Concentrations in Soil after the incorrect Burning of Contaminated Blue Lime in an Alpine Valley

Presenter: Tanja MAIER, University of Klagenfurt, Austria
Co-Authors: Gunter SPÖCK, Kurt SCHEIDL

Between 2012 and 2014, a huge amount of blue lime - contaminated with hexachlorobenzene (HCB) and mercury (Hg) - has been incorrectly burned by a cement producer in Görtschitztal, Carinthia. Consequently, huge amounts of pollutants were released in the air and got deposited on the ground. In order to analyze the environmental pollution,

samples of HCB and Hg were collected at various ground locations in the vicinity of Görtschitztal.
In this context, geostatistcal methods, like kriging, can be used to predict the concentration of HCB and Hg at unknown (not yet sampled) locations, resulting in so-called prediction maps. The spatial prediction method of kriging has two crucial assumptions. Firstly, the underlying random field has to be Gaussian. Secondly, the covariance function has to be known.
Unfortunately, these assumptions do not hold for the sampled HCB and Hg data and traditional kriging methods cannot be used as they will underestimate the prediction error. Due to this fact, trans-Gaussian kriging was used to analyze the data. This method can be used, in case the underlying random field can be appropriately transformed to a Gaussian random field. To be precise, Box-Cox-transformation was used on the data to ensure normality. Furthermore, the covariance function, which was assumed to be known, was estimated from the empirical data. Ordinary kriging was then applied to the transformed data. The estimated HCB and Hg concentrations at unsampled locations were then backtransformed to construct prediction maps. These maps indicate - and further analyzes showed - that concentrations are high in the whole valley and very high especially near Brückl and Wietersdorf, two villages in Görtschitztal.

ESA1: Ergodicity and sensitivity analysis of stochastic systems and networks I
Auditorium Maximum (HS 401)

4:00pm - 4:25pm
On Ergodicity Problem Of Some Queuing Process For Networks
Presenter: Elmira Yu. KALIMULINA, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Russian Federation
We will consider some model describing unreliable queueing network. One of the key elements of queueing network is a routing matrix, which describes the graph of our network. Its a stochastic matrix supposed to be fixed usually. Several standard algorithms of rerouting are used in the case of unreliable network nodes. We use a different approach. We consider the sequences of failure and recovery rates for our nodes. Instead of a fixed graph of our network the set of graphs and the sequence of co-matrixes describing the way of transformation of one graph from this set to another are considered. The influence of degree distribution on stability of the network is discussed and the ergodicity conditions for this model and the rate of convergence to the stationary distribution are obtained.

4:25pm - 4:50pm
On Some Generalization Of Lorden's Inequality For Renewal Processes
Presenter: Elmira Yu. KALIMULINA, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Russian Federation
Co-Authors: Galina A. ZVERKINA
The queueing system described by renewal process is considered. Two random variables called backward renewal time and forward renewal time for this process are dened. The

Lorden's inequality its true for renewal process, so expectations of backward and forward renewal times are bounded by the relation of expectation of moment of random variable X for any fixed moment of time, where X - i.i.d. We generalised and proved a similar result for dependent random variables with finite expectations, some constant C and integrable function $Q(s)$: if X are not independent and have absolutely continuous distribution function which satisfies some boundary conditions, then the analogue of Lorden's inequality for renewal process is true.

4:50pm - 5:15pm

Performance Analysis Of Heterogeneous Queueing System With Random Capacities

Presenter: Mais P. FARHADOV, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Russian Federation

In this paper a queueing system with unlimited number of servers of two different types is considered. Customers arrive in the system according to a Markov Modulated Poisson Process (MMPP) and each customer carries a random quantity of work (capacity of the customer). It is shown that the joint probability distribution of the customers number and capacities in the system has the form of Gaussian multidimensional probability distribution under the asymptotic condition of equivalently growing service time on devices of dierent type.

FTS1: Functional Time Series
Blue Lecture Hall (HS 402)

4:00pm - 4:25pm

Testing for Hidden Periodicities in Functional Time Series

Presenter: Vaidotas CHARACIEJUS, Département de mathématique, Université libre de Bruxelles, Belgium

Co-Authors: Clément CEROVECKI, Siegfried HÖRMANN

We propose several procedures to test for the presence of periodicities in functional time series when the length of the period is unknown. The tests are based on the asymptotic distribution of the maximum over all Fourier frequencies of the Hilbert-Schmidt norm of the periodogram operator of independent and identically distributed random elements with values in a real separable Hilbert space. Our approach is based on a projection onto a finite dimensional subspace spanned by a finite number of principal components. When the number of principal components is fixed, we show that the maximum converges in distribution to the standard Gumbel distribution as the sample size increases. Under stronger assumptions, we show that the same limit holds even if we let the number of principal components grow to infinity as the sample size increases. We use our asymptotic results to propose several tests for hidden periodicities in functional time series and illustrate their performance using a Monte Carlo simulation study.

4:25pm - 4:50pm

Reconstructing Partially Observed Functional Data with (Non-)Systematically Missing Parts

Presenter: Dominik LIEBL, University Bonn, Germany
Co-Authors: Alois KNEIP

The first part of the talk considers the case of partially observed functional data with non-systematically missing parts. A new reconstruction operator is proposed which aims to recover the missing parts of a function given the observed parts. This new operator belongs to a new, very large class of functional operators which includes the classical regression operators as a special case. The optimality of our reconstruction operator is shown and it is demonstrated that the usually considered regression operators generally cannot be optimal reconstruction operators. The estimation theory allows for autocorrelated functional data and considers the practically relevant situation in which each of the n functions is observed at m_i, $i = 1, \ldots, n$, discretization points plus noise. Rates of consistency are derived for the nonparametric estimation procedures using a double asymptotic. The second part of the talk proposes new estimators for the mean and the covariance function for partially observed functional data using a detour via the fundamental theorem of calculus. These new estimators allow for consistent estimation of the mean and covariance function under specific violations of the missing-completely-at-random assumption.

4:50pm - 5:15pm

Inference for the Autocovariance of a Functional Time Series, and Goodness-of-Fit Tests for FGARCH Models

Presenter: Gregory RICE, Universtiy of Waterloo
Co-Authors: Piotr KOKOSZKA, Hanlin SHANG, Yuqian ZHAO, Tony WIRIJANTO

Most methods for analyzing functional time series rely on the estimation of lagged autocovariance operators or surfaces. Testing whether or not such operators are zero is an important diagnostic step that is well understood when the data, or model residuals, form a strong white noise. When functional data are constructed from dense records of, for example, asset prices or returns, a weak white noise model allowing for conditional heteroscedasticity is often more realistic. Applying inferential procedures for the autocovariance based on a strong white noise to such data often leads to the erroneous conclusion that the data exhibit significant autocorrelation. We develop methods for performing inference for the lagged autocovariance operators of stationary functional time series that are valid under general conditional heteroscedasticity conditions, and apply these to conduct goodness-of-fit tests for fGARCH models.

SL2: Statistical Learning - Methods and Applications II
Green Lecture Hall (HS 403)

5:15pm - 5:40pm

Clustering and Symptom Analysis in Binary Data with Application

Presenter: Fatema Sadik AL-JUBOORI, St. Petersburg State University, Russian Federation

Co-Authors: Nina Petrovna ALEKSEYEVA, Evgeniia Petrovna SKURAT
The Canonical Super Symptom Analysis can be used to identify the structure of interrelations between two sets of categorical variables on the base of polinomial combinations over the field F2. In oncology, this method allows to select a complex component factor for predicting forms of breast cancer. The problem is the complexity of busting. To reduce the enumeration of all the options, it is proposed to apply super symptoms to describe the clusters as a subsets of multidimensional space over F2.

5:40pm - 6:05pm
Secondary Use of Clinical Problem List Entries for Data-Driven Learning Approaches
Presenter: Markus KREUZTHALER, Medical University of Graz, Austria
Co-Authors: Michel OLEYNIK, Jose Antonio VERA RAMOS, Zdenko KASAC, Stefan SCHULZ
Supervised learning systems depend on large amounts of annotated data. The annotation effort of manual data decoration contradicts this requirement for applied neural networks. Clinical information systems are a rich resource of patient-level information, partially annotated with administrative codes for disorders and procedures. This makes them interesting for data-driven learning approaches. We investigated to what extent concise (50 character) clinical problem list entries can be used to automatically assign codes from the International Classification of Diseases (ICD-10) to each of them. The input data was represented as a sequential time series of one-hot encoded single characters for the proposed stacked (BI-)LSTM neural network architecture. As a baseline we used fastText from Facebook's AI Research (FAIR) Lab. Our proposed model equalizes fastText with a macro-averaged F-measure of 0.83 tested on the top 100 most frequent ICD-10 codes. A detailed error inspection in combination with a neural network activation analysis has shown inconsistent manual coding as upper performance bound.

6:05pm - 6:30pm
Extraction of Poorly Visible Grain Boundaries from Tomographic Image Data, Using Convolutional Neural Networks
Presenter: Orkun FURAT, Institute of Stochastics, Ulm University, Germany
Co-Authors: Mingyan WANG, Matthias NEUMANN, Lukas PETRICH, Matthias WEBER, Carl E KRILL III, Volker SCHMIDT
The investigation of the morphology of grain boundaries in polycrystalline materials is of interest in the field of materials science. Sophisticated measurement techniques, like 3D X-ray diffraction (3DXRD), can determine crystallographic orientations of grains and thus provide the location of grain boundaries. However, such methods are expensive, time consuming and difficult to perform in situ, e.g., during thermodynamic treatment. More available techniques, such as X-ray microtomography, often provide challenging data when imaging alloys, since grain boundaries do not induce contrast in computed tomography (CT) image data. This work deals with image data of an aluminum-copper

specimen in which a liquid attaches to grain boundaries during Ostwald ripening. Since the contrast between grain interior and boundaries can be poor in CT data for low amounts of liquid, it is difficult to extract the grain boundaries from CT data with conventional image processing techniques. Therefore, a convolutional neural network was trained with matching pairs of CT and 3DXRD data to detect poorly visible grain boundaries solely from CT data. The network's output was then segmented into grains with conventional image processing techniques like the watershed transform. This approach leads to a sufficiently good segmentation of grains in CT data such that quantitative analysis and stochastic modeling of the grain microstructure are possible.

SSP1: Statistical Selection Procedures and Multiple Comparison
Blue Lecture Hall (HS 402)

5:15pm - 5:40pm

Controlling The False Discovery Rate In Linear And Logistic Regression With SLOPE

Presenter: Michał KOS, University of Wrocław, Poland
Co-Authors: Małgorzata BOGDAN

Sorted L-One Penalized Estimator (SLOPE) is a solution to a following convex optimization problem:

$$b^{est} = argmin_b[-l(b) + \sum_i \lambda_i |b|_{(i)}]$$

where: $b = (b_1, \ldots, b_p)'$ is a vector, $l(b)$ is a loglikelihood function of linear or logistic regression, the $\lambda_1, \lambda_2, \ldots, \lambda_p$ is a positive, non-increasing sequence of tuning parameters and the $|b|_{(i)}$ is the i-th largest element of $(|b_1|, \ldots, |b_p|)'$. SLOPE chooses columns of the design matrix $X_{n \times p}$ that are associated with non-zero elements of b^{est} and identifies them as important predictors. In linear regression, when the design matrix X is orthogonal, SLOPE with the sequence of tuning parameters selected according to the thresholds of the Benjamini - Hochberg (BH) procedure for multiple testing controls the False Discovery Rate (FDR).

During the session we will present new results illustrating that SLOPE asymptotically controls FDR for the linear and logistic regression, when entries of the design matrix are iid variables from the normal distribution. We will discuss both low dimensional set-up, where 'p' is fixed and 'n' goes to infinity, and the high dimensional set-up, where 'p' may diverge to infinity much quicker than 'n'. We will illustrate our asymptotic results with computer simulations. Apart from the Gaussian design matrix we will also consider the practical case of the design matrix containing genotypes of independent genetic markers.

Reference:
1. M. Bogdan, E. van den Berg, C. Sabatti, W. Su, E. Candes (2015). SLOPE - Adaptive Variable Selection via Convex Optimization. Annals of Applied Statistics Vol. 9, No. 3, 1103-1140

5:40pm - 6:05pm

Approximately Optimal Subset Selection for Statistical Design and Modelling

Presenter: James V ZIDEK, University of British Columbia, Canada
Co-Authors: Yu WANG, Nhu D LE
We study the problem of optimal subset selection from a set of correlated random variables. In particular, we consider the associated combinatorial optimization problem of maximizing the determinant of a symmetric positive definite matrix that characterizes the chosen subset. This problem arises in many domains, such as experimental designs, regression modelling, and environmental statistics. We establish an efficient polynomial-time algorithm using the determinantal point process to approximate the optimal solution to the problem. We demonstrate the advantages of our methods by presenting computational results for both synthetic and real data sets.

6:05pm - 6:30pm
Asymptotic Confidence Regions Based On The Adaptive Lasso With Partial Consistent Tuning
Presenter: Nicolai David AMANN, University of Vienna, Austria
Co-Authors: Ulrike SCHNEIDER
We construct confidence sets based on an adaptive Lasso estimator with componentwise tuning in the framework of a low-dimensional linear regression model. We consider the case where at least one of the components is penalized at the rate of consistent model selection and where certain components may not be penalized at all. We perform a detailed study of the consistency properties and the asymptotic distribution that includes the effects of componentwise tuning within a so-called moving-parameter framework. These results enable us to explicitly provide a set M such that every open superset acts as a confidence set with uniform asymptotic coverage equal to 1 whereas every proper closed subset with non-empty interior is a confidence set with uniform asymptotic coverage equal to 0. The shape of the set M depends on the regressor matrix as well as the deviations within the componentwise tuning parameters.

Thursday September 5, 2019

LF1: Likelihood-free Statistical Design and Inference
Green Lecture Hall (HS 403)

10:30am - 10:55am
Optimal Bayesian Design For Models With Intractable Likelihoods Via Supervised Learning Methods
Presenter: Markus HAINY, Johannes Kepler University, Austria
Co-Authors: David James PRICE, Olivier RESTIF, Christopher DROVANDI
Optimal Bayesian experimental design is often computationally intensive due to the need to approximate many posterior distributions for datasets simulated from the prior predictive distribution. The issues are compounded further when the statistical models of interest do not possess tractable likelihood functions and only simulation is feasible. We

employ supervised learning methods to facilitate the computation of utility values in optimal Bayesian design. This approach requires considerably fewer simulations from the candidate models than previous approaches using approximate Bayesian computation. The approach is particularly useful in the presence of models with intractable likelihoods but can also provide computational advantages when the likelihoods are manageable. We consider the two experimental goals of model discrimination and parameter estimation. The methods are applied to find optimal designs for models in epidemiology and cell biology.

10:55am - 11:20am

Bayesian Design For Intractable Likelihood Models

Presenter: Antony OVERSTALL, University of Southampton, United Kingdom

Bayesian designs are found by maximising the expectation of a utility function where the utility function is chosen to represent the aim of the experiment. There are several hurdles to overcome when considering Bayesian design for intractable models. Firstly, common to nearly all Bayesian design problems, the expected utility function is not analytically tractable and requires approximation. Secondly, this approximate expected utility needs to be maximised over a potentially high-dimensional design space. To compound these problems, thirdly, the likelihood is intractable, i.e. has no closed form. New approaches to maximise an approximation to the expected utility for intractable models are developed and applied to illustrative exemplar design problems with experimental aims of parameter estimation and model selection.

11:20am - 11:45am

Efficient Bayesian Experimental Design for Implicit Models

Presenter: Steven KLEINEGESSE, University of Edinburgh, United Kingdom
Co-Authors: Michael GUTMANN

Bayesian experimental design involves the optimal allocation of resources in an experiment, with the aim of optimising cost and performance. For implicit models, where the likelihood is intractable but sampling from the model is possible, this task is particularly difficult and therefore largely unexplored. This is mainly due to technical difficulties associated with approximating posterior distributions and utility functions. We devise a novel experimental design framework for implicit models that improves upon previous work in two ways. First, we use the mutual information between parameters and data as the utility function, which has previously not been feasible. We achieve this by utilising Likelihood-Free Inference by Ratio Estimation (LFIRE) to approximate posterior distributions, instead of the traditional approximate Bayesian computation or synthetic likelihood methods. Secondly, we use Bayesian optimisation in order to solve the optimal design problem, as opposed to the typically used grid search or sampling-based methods. We find that this increases efficiency and allows us to consider higher design dimensions.

11:45am - 12:10pm
Invited Discussion of the Talks
Presenter: Christian P ROBERT, University Paris Dauphine, France
Co-Authors: Jürgen PILZ
Christian Robert and Jürgen Pilz will discuss the talks that have been presented in this session.

SSM1: Stochastic Modelling and Simulation in Materials Science and Engineering I
Blue Lecture Hall (HS 402)

10:30am - 10:55am
Joint Distribution of Local Porosity and Local Tortuosity in Sack Paper
Presenter: Eduardo MACHADO-CHARRY, Graz University of Technology, Austria
Co-Authors: Matthias NEUMANN, Peter LEITL, Ulrich HIRN, Volker SCHMIDT, Karin ZOJER
The porosity of paper is a crucial quantity for most of its applications. The statistical analysis of the 3D porous microstructure bears enormous benefits for accessing the role of porosity in determining strongly related properties such as the air permeance of Sack paper: The local variations in porosity are crucial as pores are nonuniformly distributed across a paper sheet. The air permeance is not solely governed by the porosity, but e.g., also the sineousness of transport paths, measured by the mean geodesic tortuosity. We determine the local variability of porosity in sack paper and the local mean geodesic tortuosity from multiple, non-overlapping cutouts of the microstructure. The distribution of these quantities shows a negative correlation between local porosity and tortuosity. Using a copula approach, we model the corresponding joint probability distribution to quantify the porosity-tortuosity correlation. The approach allows, e.g., for a prediction of the conditional distribution of local mean geodesic tortuosities for a given value of local porosity.

10:55am - 11:20am
Imaging, Simulation, and Characterization of Porous Materials and Their Mass Transport Properties
Presenter: Magnus RÖDING, RISE Research Institutes of Sweden AB, Sweden
To understand porous material microstructures and their mass transport properties, and ultimately learn how to tailor materials for specific purposes, quantitative characterization using statistical models and methods is key. We will discuss a number of cases involving image analysis of microstructures, generation of realistic virtual materials, and mass transport simulation and prediction from microstructural geometry. In image analysis, we will cover image segmentation of focused ion beam scanning electron microscopy (FIB-SEM) 3D data from porous polymer films and paperboard coating layers to extract microstructural information. For generation of virtual materials structures, we will give examples of different morphologies such as heterogeneous fiber materials, nanoplatelet-filled composites, ordered and disordered granular materials, and porous structures simulated using

phase-separation dynamics and thresholded Gaussian random field models. We will discuss simulation of effective diffusivity and fluid permeability in these virtual structures, and how different microstructural descriptors such as tortuosity and correlation functions can be used, either in regression or machine learning models, to predict and understand the relationships between microstructure and mass transport properties.

11:20am - 11:45am

Reconstruction and Stochastic 3D Modeling of Grain Boundaries in Polycrystalline Materials from Incomplete Data, Using Tessellations with Curved Facets

Presenter: Lukas PETRICH, Institute of Stochastics, Ulm University, 89069 Ulm, Germany
Co-Authors: Daniel WESTHOFF, Jakub STANĚK, Mingyan WANG, Carl E. KRILL III, Volker SCHMIDT

The curvature of grain boundaries in polycrystalline materials is an important characteristic since it plays a key role in phenomena such as grain growth. However, most traditional tessellation models that are used for stochastic modeling of these materials, e.g. Voronoi or Laguerre tessellations, have flat faces, and thus fail to incorporate the curvature.
For this reason, we present a stochastic model for polycrystalline AlCu samples based on spherical growth tessellations, which exhibit non-convex cells. The generators of these tessellations can be thought of as the midpoints and radii of spheres. The core of the stochastic model is a collective rearrangement algorithm that shifts these spheres in such a way that the resulting tessellation statistically matches selected functional characteristics of the material samples.
This calibration step can be performed on full information of the grain boundaries, which, for example, can be acquired by three-dimensional electron backscatter diffraction (3D-EBSD). However, we also present a method to extract the necessary characteristics from far-field three-dimensional X-ray diffraction (3D-XRD) microscopy. The advantage of this measurement technique is that it allows to quickly investigate large numbers of grains without destroying the specimen in the process, rendering it a good solution for in situ experiments. However, this comes at the cost of only being able to collect aggregated morphological information, i.e., the centers of mass and the volumes of the grains.

11:45am - 12:10pm

3D Optical Flow for the Estimation of Motion Vector Fields

Presenter: Tessa KUSCHNERUS, Fraunhofer ITWM and University Kaiserslautern, Germany
Co-Authors: Claudia REDENBACH, Katja SCHLADITZ

Estimating motion vector fields from in-situ testing is one of the modern challenges in 3D image analysis. Experimental setups that combine scanning and loading devices can give completely new insight into the formation of cracks or failure for material such as concrete or ceramic foams. Unfortunately, these setups produce large datasets and the efficient

exploration of motion from these is still an open question. Many reliable algorithms in two dimensions are not applicable in 3D as plain extension is too costly. We propose an efficient extension to robust optical flow algorithms and compare them to the state-of-art motion estimation by using Digital Volume Correlation (DVC) and CLAIRE (Constrained Large Deformation Diffeomorphic Image Registration). Robustness in these algorithms is achieved by additional regularizers that exploit dense feature sets and by integrating a coarse-to-fine scheme in the solution process. The optimal solution is obtained by minimizing a functional. As in 2D optical flow, we present a solution for a linearized, convex formulation and for a variational approach without linearization.

CB1: Algebraic Methods in Computational Biology
Blue Lecture Hall (HS 402)

1:00pm - 1:25pm
Statistics on Stratified Spheres for Analysis of RNA structure
Presenter: Stephan HUCKEMANN, University of Göttingen, Germany
Co-Authors: Benjamin ELTZNER, Kanti MARDIA
It is highly popular to model RNA molecules via dihedral angles, leading to a data on a torus. We will see that this is very unfavorable towards PCA-like dimension reduction methods. Keeping the topology (almost) and suitably changing the geometry leads to a stratified sphere, which turns out to be most favorable towards PCA-like methods.

1:25pm - 1:50pm
Information Geometry for Phylogenetic Trees
Presenter: Tom NYE, Newcastle University, United Kingdom
Co-Authors: Maryam Kashia GARBA
Phylogenetic trees, which represent evolutionary relationships between present-day species, are usually inferred from gemetic sequence data. As such, each tree represents a different Markov model of sequence evolution. Collections or samples of alternative trees arise as a result of uncertainty when inferring trees from data. In recent years a number of powerful geometric methods have been developed for analysing samples of phylogenetic trees, for tasks such as computing sample means and performing principal component analysis. Typically, these methods adopt certain underlying geometric assumptions which regard trees as purely geometric objects in terms of branching shape and edge lengths. We develop an alternative approach which constructs a geometry on tree space by regarding trees directly as probabilistic models of sequence evolution. First we describe certain metrics on tree space which are induced by metrics between distributions, such as the Hellinger distance. For large trees, a simulation strategy must be adopted to compute approximate distances. These metrics show radically different behaviour from existing metrics between trees. Secondly, we compute geodesics in tree space using the Riemannian metric defined by the Fisher information matrix. These geodesics are computationally expensive to construct, but we show via examples how they are closely approximated by geodesics in a more tractable geometry obtained by embedding tree space in the space of covariance matrices. These metrics and the associated geodesics open up a number of new possibilities

for analysing samples of phylogenetic trees.

1:50pm - 2:15pm
Trait Evolution On Two Gene Trees
Presenter: James DEGNAN, University of New Mexico, United States of America
Models of trait evolution use a phylogenetic tree to determine the correlation structure for traits sampled from a set of species. Typically, the phylogenetic tree is estimated from genetic data from many loci, and a single tree is used to model the trait evolution, for example by assuming that the mean trait value follows a Brownian motion on the tree. Here, we model trait evolution by assuming that there are two genetic loci influencing the trait. In this case separate evolutionary trees (called gene trees) can occur for the two loci. We model the correlation structure as arising from a linear combination of Brownian motions on the two trees, and develop a model to estimate the proportion of trait evolution contributed by each gene.

MNT1: Modern Statistics and New Technologies
Green Lecture Hall (HS 403)

1:00pm - 1:25pm
Deep Bayesian Regression
Presenter: Florian FROMMLET, Medical University Vienna, Austria
One of the most exciting recent developments in data analysis is deep learning. Multilayer networks have become extremely successful in performing prediction tasks and are successfully applied in many different areas. However, the resulting prediction models often difficult to interpret and potentially suffer from overfitting. The aim of this paper is to bring the ideas of deep learning into a statistical framework which yields more parsimonious models and allows to quantify model uncertainty. To this end we introduce the class of deep Bayesian regression models (DBRM) consisting of a generalized linear model combined with a comprehensive non-linear feature space, where non-linear features are generated just like in deep learning. DBRM can easily be extended to include latent Gaussian variables to model complex correlation structures between observations, which seems to be not easily possible with existing deep learning approaches. Two different algorithms based on MCMC are introduced to fit DBRM and to perform Bayesian inference. The predictive performance of these algorithms is compared with a large number of state of the art learning algorithms. Furthermore we illustrate how DBRM can be used for model inference in various applications.

1:25pm - 1:50pm
Flexible Bayesian approaches for genomics data
Presenter: Francesco Claudio STINGO, University of Florence, Italy
In this talk, I will describe recent approaches for the analysis of multi-omics data. Bayesian integration of heterogeneous data can facilitate the identification of patient-specific prognostic biomarkers, a critical step in the development of personalized treatment for clin-

ically and molecularly heterogeneous diseases such as cancer. The proposed methods allow flexible modeling of the biological interactions, as well as induces sparsity resulting in more parsimonious and interpretable models. Simulation studies demonstrate the superior performance of the proposed methods against competing method in terms of both marker selection and prediction. The application of the proposed methodology results in a better understanding of the underlying biological mechanisms.

1:50pm - 2:15pm

Fast Nonparametric Bootstrap for Multivariate Time Series

Presenter: Michele LA ROCCA, University of Salerno, Italy
Co-Authors: Cira PERNA

The talk aims to propose and discuss a sieve bootstrap scheme based on Extreme Learning Machines for multivariate time series. In ELM modeling, the learning process can be done without iteratively tuning the hidden nodes (randomly choosing the hidden layer weights) while retaining the property of being universal approximators for general nonlinear functions, without imposing any restrictive assumption on the activation function. Therefore, using ELMs in the resampling scheme can dramatically reduce the computational burden of the overall bootstrap procedure, with performances comparable to the Neural Network Sieve bootstrap and computing time similar to the AR-Sieve bootstrap. Moreover, ELMs are well suited for massive data processing and so the novel bootstrap scheme is expected to scale up nicely for high-dimensional time series, making this tool suitable for inference on complex time series vectors. The overall proposed bootstrap procedure has the advantage of being fully nonparametric while retaining the conceptual simplicity of the residual bootstrap.

BDA1: Big Data Analytics and High-Dimensional Data Analysis
Auditorium Maximum (HS 401)

2:15pm - 2:40pm

Multidimensional Sequential Pattern to Find Causes of Problems

Presenter: Zornica Vaskova VASILEVA, Liebherr-Werk Nenzing GmbH, Austria
Co-Authors: Christian BORGELT

Nowadays machines are like a computer and have many sensors. Thus, each machine generates logs of items. An item contains, for example, information, warning or error data. The target is to find machines with unknown problems or unexpected behaviours comparing the data of a group of similar machines. Then again, if a certain problem or failure has occurred, it is very interesting to find the reason. It should be noted that the number of items increases constantly, so the data changes over time. Therefore, we should use a quick and efficient method to analyse our machine data. The methods for pattern mining, used by the market basket analysis, can be adapted to analyse machine data. A log item can be seen as a product in a supermarket and a machine can be seen as a customer. In machine data analysis, the order of log items is very important. Therefore, we are interested in sequences of items, which occur frequently in machine data. Thus, we search for (maximal) sequential pattern in machine data. In this study, we design a method that

finds similar sequential patterns per group of machines. Additionally, we use this method to find the reason for a known error item or a certain failure in a group of machines that share this error. We applied our new method extensively on known problems and certain failures. The method performed remarkably well. It not only found the expected results, in the known cases, but furthermore detected valuable, previously unknown information about the machines. These newly found patterns can now be matched in machine data of different machine types. Analysing the machine data according to the introduced method proved to be very beneficial.

2:40pm - 3:05pm

Asymptotic Distribution Of Test Statistic For Two Sample Test Under High-Dimensional Setting

Presenter: Takahiro NISHIYAMA, Senshu university, Japan
Co-Authors: Masashi HYODO, Tatjana PAVLENKO

We discuss the canonical testing problem in modern statistical inference, namely the two-sample test for equality of mean vectors of independent multivariate normal populations with large dimensions. For two-sample test, L_2 type test statistic was proposed by Chen and Qin (2010, Ann. Statist.). In this talk, we derive two new asymptotic approximations for a distribution of this L_2 type test statistic. Also, we establish explicit error bounds of these two new asymptotic approximations. Finally, we verify the accuracy of the proposed approximations via Monte Carlo simulations.

3:05pm - 3:30pm

Robust Bayesian Inference using γ-divergence

Presenter: Tomoyuki NAKAGAWA, Tokyo University of Science, Japan
Co-Authors: Shintaro HASHIMOTO

In Bayesian analysis, it is well known that ordinary Bayesian estimator is not robust against outliers. Ghosh and Basu (2016) proposed the robust Bayesian estimation against outliers by using the density power divergence. They characterized the robustness in terms of the influence function. However, in frequentist viewpoint, an estimator using the density power divergence does not work well the estimation for the scale parameter, and unstable when the contamination ratio is not small. On the other hand, it was shown that an estimator using the γ-divergence can make a stable estimation even when the contamination ratio is not small. Therefore, in the presentation, we propose a robust estimation using γ-divergence. Furthermore, The accuracy of the estimators depends on the prior distribution. Thus, we also propose the moment matching priors for the robust Bayesian estimation, and illustrate its performances in some simulation studies.

3:30pm - 3:55pm

Problems Of Applying Tests Of Checking Statistical Hypotheses When Analyzing Big Data

Presenter: Boris Yu. LEMESHKO, Novosibirsk State Technical University, Russian Federation
Co-Authors: Stanislav B. LEMESHKO
When analyzing large samples, there are problems that limit the application of classical results. The paper discusses methods for constructing estimates in the analysis of big data. The effect on the conclusions by the χ^2 Pearson test of the choice of the number of intervals and the method of grouping is demonstrated.
Using statistical modeling methods, it was shown that the main reason for incorrect conclusions when using for analysis of large samples of different tests for checking hypotheses is the limited accuracy of data presentation. This applies to the set of non-parametric goodness-of-fit tests (Kolmogorov, Kramer – Mises – Smirnov, Anderson – Darling, etc.), as well as many special tests for checking normality, uniformity, etc.
It is shown that the distribution of statistics of the homogeneity tests of laws, as well as the homogeneity tests of averages and homogeneity tests of variances is influenced by the unevenness of data representation in compared samples.
Recommendations are given on the application of tests for analyzing large samples, and software is offered to solve problems.

DGA1: Data generation assisted inference
Blue Lecture Hall (HS 401)

2:15pm - 2:40pm
Causal Effects on Networks and cluster data under Interference
Presenter: Fabrizia MEALLI, University of Florence, Italy
In most real-world systems units are interconnected and can be represented as networks consisting of nodes and edges; clusters are special types of networks. In settings where some units are exposed to a treatment and its effects spills over connected units, estimating both the direct effect of the treatment and spillover effects presents several challenges. First, assumptions on the way and the extent to which spillover effects occur along the observed network are required. Second, in observational studies, where the treatment assignment is not under the control of the investigator, confounding and homophily are potential threats to the identification and estimation of causal effects on networks. Under neighborhood interference and unconfoundedness of the individual and neighborhood treatment we develop a new covariate-adjustment estimator for treatment and spillover effects in observational studies on networks. Estimation is based on a generalized propensity score that balances individual and neighborhood covariates across units under different levels of individual treatment and of exposure to neighbors' treatment. Correlation of interacting units is taken into account using a community detection algorithm and incorporating random effects in the outcome model.

2:40pm - 3:05pm
Community Detection with Dependent Connectivity
Presenter: Annie QU, UIUC, United States of America

In network analysis, within-community members are more likely to be connected than between-community members, which is reflected in that the edges within a community are intercorrelated. However, existing probabilistic models for community detection such as the stochastic block model (SBM) are not designed to capture the dependence among edges. In this paper, we propose a new community detection approach to incorporate within-community dependence of connectivities through the Bahadur representation. The proposed method does not require specifying the likelihood function, which could be intractable for correlated binary connectivities. In addition, the proposed method allows for heterogeneity among edges between different communities. In theory, we show that incorporating correlation information can lower estimation bias and accelerate algorithm convergence. Our simulation studies show that the proposed algorithm outperforms the popular variational EM algorithm assuming conditional independence among edges. We also demonstrate the application of the proposed method to agricultural product trading networks from different countries. This is joint work with Yubai Yuan of UIUC.

3:05pm - 3:30pm

Testing for the Rank of a Covariance Operator by Matrix Completion

Presenter: Victor PANARETOS, EPFL, Switzerland

How can we discern whether a continuous-time stochastic process is finite-dimensional, and if so, what its precise dimension is? And how can we do so at a given level of confidence? This question is central to a great deal of methods for functional data, which require low-dimensional representations whether by functional PCA or other methods. The difficulty is that the determination is to be made on the basis of iid replications of the process observed discretely and with measurement error contamination. This adds a ridge to the empirical covariance, obfuscating the underlying dimension. We build a matrix-completion inspired test statistic that circumvents this issue by measuring the best possible least square fit of the empirical covariance's off-diagonal elements, optimised over covariances of given finite rank. For a fixed grid of sufficient size, we determine the statistic's asymptotic null distribution as the number of replications grows. We then use it to construct a bootstrap implementation of a stepwise testing procedure controlling the family-wise error rate corresponding to the collection of hypotheses formalising the question at hand. Under minimal regularity assumptions we prove that the procedure is consistent and that its bootstrap implementation is valid. The procedure involves no tuning parameters or pre-smoothing, is indifferent to the homoskedasticity or lack of it in the measurement errors, and does not assume a low-noise regime. Based on joint work with Anirvan Chakraborty (IISER Calcutta)

3:30pm - 3:55pm

The Five Trolls under the Bridge: Principal Component Analysis with Asynchronous and Noisy High Frequency Data

Presenter: Per Aslak MYKLAND, University of Chicago, United States of America

We develop a principal component analysis (PCA) for high frequency data. As in Northern fairly tales, there are trolls waiting for the explorer. The first three trolls are market microstructure noise, asynchronous sampling times, and edge effects in estimators. To get around these, a robust estimator of the spot covariance matrix is developed based on the Smoothed TSRV. The fourth troll is how to pass from estimated time-varying covariance matrix to PCA. Under finite dimensionality, we develop this methodology through the estimation of realized spectral functions. Rates of convergence and central limit theory, as well as an estimator of standard error, are established. The fifth troll is high dimension on top of high frequency, where we also develop PCA. With the help of a new identity concerning the spot principal orthogonal complement, the high-dimensional rates of convergence have been studied after eliminating several strong assumptions in classical PCA. As an application, we show that our first principal component (PC) closely matches but potentially outperforms the S&P 100 market index. From a statistical standpoint, the close match between the first PC and the market index also corroborates this PCA procedure and the underlying S-TSRV matrix, in the sense of Karl Popper. This is joint work with Dachuan Chen and Lan Zhang.

SMR1: Statistical methods and applications in medical research
Green Lecture Hall (HS 403)

2:15pm - 2:40pm

Data-generating Mechanism And Simulation For Random-effects Network Meta-analysis Of Binary Outcome Including Multi-arm Trials

Presenter: Svenja SEIDE, Institute of Medical Biometry and Informatics, University of Heidelberg, Germany
Co-Authors: Katrin JENSEN, Meinhard KIESER

Statistical methods are commonly evaluated by means of simulation studies. In case of network meta-analysis of binary data however, available data-generating models are restricted to either inclusion of two-armed trials, or the common-effect (also called fixed-effect) model. We evaluate a possible extension of the data-generating mechanisms for random-effects network meta-analysis including multi-arm trials based on data-generation in the pairwise case recently discussed by Pateras *et al.* (2018). As only one of the data-generating models discussed for the pairwise case is directly applicable in a (random-effects) network setting, and this data-generating model uses strongly restrictive assumptions, we propose a modified version for the use in simulations studies. Based on this modified data-generating model, a simulation procedure for the simulation of a network of treatments using ORs as effect measure and including multi-arm trials in a random-effects setting, is proposed and its performance evaluated on synthetic data and in an empirical example which is adapted from Warren *et al.* (2014).

References:

[1] Pateras K, Nikolakopoulos S, Roes K. Data generating models of dichotomous outcomes: Heterogeneity in simulation studies for a random-effects meta-analysis. Statistics in Medicine. 2018; 37:1115–1124. https://doi.org/10.1002/sim.7569.
[2] Warren, F. C., Abrams, K. R. and Sutton, A. J. (2014), Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with

regard to adverse outcomes. Statist. Med., 33: 2449–2466. doi:10.1002/sim.6131.

2:40pm - 3:05pm

U-NET: Segmentation Of Electron Microscopic Cell Recordings

Presenter: Wolfgang HITZL, Paracelsus Medical University Salzburg, Austria

Aims: Statistical learning theory has led to many successful applications in semantic segmentation of images. We illustrate U-NET, a CNN architecture for fast and precise segmentation of images which was developed at the Computer Science department at the University of Freiburg, Germany.

Methods: We apply U-NET to $n = 135$ transmission electron microscopic cell recordings to segment cell membranes and cell nuclei in a training and independent test sample. This data set is provided by the computer vision group of the University Freiburg. We illustrate a typical workflow for image segmentation - preprocessing, data augmentation, training and testing - by use of MATHEMATICA 12.

Results: The net generalized very well. In the total sample, Intersection over Union (IoU) for background, cell membrane and cell nuclei was 99% (95% CI: 98.5-99.2), 94% (95% CI: 93-95) and 92% (89-85), respectively.

Discussion: Advantages and possible limitations of U-NET will be discussed of this approach.

3:05pm - 3:30pm

Blinded Sample Size Recalculation For Ancova Models With Multiple Random Covariates

Presenter: Georg ZIMMERMANN, Department of Mathematics, Paris Lodron University Salzburg

Co-Authors: Meinhard KIESER, Arne C. BATHKE

Consider the situation of comparing two groups of patients with respect to a univariate outcome of interest, adjusting for one or several covariates. If the outcome variable is continuous, the adjusted group means are usually compared by using the analysis of covariance (ANCOVA) approach. The case of random covariates is of particular interest, because adjustments for baseline measurements are strongly recommended by regulatory agencies. Moreover, controlling for additional variables that are supposed to be correlated with the outcome could reduce the bias of the effect estimators and increase the inferential power. However, methods for sample size recalculation are only available for ANCOVA models with a single random covariate. Therefore, we consider the case of a univariate ANCOVA model with multiple random covariates and possibly unequal group sizes. At first, we derive an asymptotic sample size formula and discuss some finite-sample adjustments. Subsequently, we propose a sample size recalculation method, which is based on the re-estimation of nuisance parameters at the interim analysis, yet without unblinding. We evaluate the properties of our proposed approach in an extensive simulation study. It turns out that the recalculation method shows a similar performance as the fixed sample size calculation approaches when all parameters have been correctly specified, but outper-

forms the latter in terms of power in case of misspecifications (i.e., when some uncertainty is present in the planning phase). The average total sample sizes from the recalculation procedure exceed the fixed sample sizes by only 6 to 7 subjects, which is the small "price to pay" for the increased flexibility.

3:30pm - 3:55pm

Simulation In Evidence Synthesis - Incorporation Of Multiple Studies In Indirect Comparisons

Presenter: Dorothea WEBER, Institute of Medical Biometry and Informatics, University of Heidelberg, Germany
Co-Authors: Katrin JENSEN, Meinhard KIESER

In the last years, indirect comparisons are widely used in evidence synthesis to compare treatments based on different primary studies in cases where direct evidence is missing. In a usual situation of an indirect comparison, one is interested in the treatment effect between treatment A and C, but direct comparisons are only available for treatment A versus B and C versus B. Thus, one would like to use the available information to get an estimate for the treatment effect of A versus C. The matching-adjusted indirect comparison (MAIC), a widely spread method for indirect comparisons, is designed for using one study per treatment comparison. However, in common scenarios there are more than one study per treatment comparison available. Simulation studies are a useful tool in evidence synthesis to evaluate scenarios for which an analytic approach is not feasible. We contrast different approaches for the inclusion of multiple studies within the MAIC and the method of Bucher by a simulation study. The challenge in simulation is the data generation of individual patient data as well as aggregated data for multiple studies and different treatment arms. In case of multiple studies including varying covariate adjustments and interactions, a prespecified treatment effect for the indirect comparison, which is needed for evaluation, is not straightforward to specify. We propose a solution for simulating a predefined treatment effect and with it an approach to interprete the results of simulation studies. Furthermore, we give recommendations in the field of indirect comparisons including multiple studies. Additionally, we investigate situations where the assumptions of methods for indirect comparisons are not met, like between-study differences which makes the simulation even more complex. The results obtained from the simulation of clinically relevant situations allows to give recommendations for practical applications.

Friday September 6, 2019

SC2: Free contributions II
Green Lecture Hall (HS 403)

10:30am - 10:50am

Computer Simulation in Study Two-Sample Tests for Randomly Right-Censored Observations

Presenter: Petr Aleksandrovich FILONENKO, Novosibirsk State Technical University, Russian Federation
Co-Authors: Sergey Nikolaevich POSTOVALOV
To apply the statistical methods it is necessary to evaluate their effectiveness. For any cases it can be done by a computer simulation, e.g. by the Monte Carlo method. By repeating a statistical procedure many times, one can obtain various statistical conclusions about the advantages and disadvantages of one statistical method over another. Using this approach, we have researched two-sample tests for randomly right-censored observations. All two-sample tests identify "differences" between null and alternative hypotheses using a different measure. However, the simulation allows one to find the advantages and disadvantages of statistical tests and formulate recommendations for their usage in practice. In our work, we consider such tests characteristics like a test power and a convergence rate to the limit distribution of the test statistic. The results of extensive computer simulations are given.

10:50am - 11:10am
Interaction Between Cytokinin and Light Signaling in Arabidopsis Thaliana - Phenotypic Study and Data Analysis
Presenter: Ioannis I. SPYROGLOU, CEITEC - Masaryk University, Brno, Czech Republic
Co-Authors: Astrid JUNKER, Zuzana GELOVÁ, Thomas ALTMANN, Jan HEJÁTKO
Interaction between light and hormonal (cytokinin) signaling is known for decades. However, the underlying molecular mechanisms are just emerging. The possible role of light/cytokinin signaling crosstalk during drought stress response was assayed in Arabidopsis thaliana via automated non-invasive phenotyping.
Data from 229 plants grown under both control (well watered) and water stress conditions are processed using the Random Forest classifier and then the results are visualized by applying Principal Component Analysis to the most significant traits. In addition, Linear Mixed Models are used to estimate adjusted means of the significant traits based on position variables to determine if there are alterations to the clustering of signaling components.
Based on the applied methods, overexpressor of cytokinin signaling component AHP3OE, light signaling mutant phyB and interaction ARR16OE/phyB have almost unique phenotypic traits. Furthermore, all the cytokinin signaling (ahp) mutants are clustered together. Importantly, the unique phenotype of AHP3OE is dependent on functional light signaling via phyA. In addition, the estimated means from Linear mixed models show that the effect of position variables on the most significant phenotypic traits derived from the Random Forest do not affect the clustering of the signaling components.
Supported by CZ.02.1.01/0.0/0.0/16_026/0008446.

11:10am - 11:30am
Multivariate Kernel Density Estimator From Ranked Set Samples

Presenter: Hikaru YAMAGUCHI, Tokyo University of Science, Japan
Co-Authors: Hidetoshi MURAKAMI
Ranked set sampling (RSS) is applicable practical methods when the variable of interest for an observed item is costly or time-consuming but the ranking of a set of items according to the variable can be easily done without actual measurement. Kernel density estimation for the multivariate data is one of nonparametric probability density estimation methods and is an important technique that has a wide range of application in econometrics and finance. In addition, multivariate kernel density estimation is used estimation of the location of the mode of a multivariate density function. We propose a method for multivariate density estimation using RSS data for multivariate samples and derive the properties of the resulted multivariate kernel density estimate. By a simulation study, it is shown that the multivariate kernel density estimate using RSS data performs better than its counterpart based on simple random sampling data by in the sense that it has smaller mean integrated square error.

11:30am - 11:50am
Numerical Models Of The Sea Surface Undulation Based On Observations
Presenter: Kristina V. LITVENKO, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of RAS, Russian Federation
Co-Authors: Sergei M. PRIGARIN
In this paper, we develop stochastic algorithms to simulate the sea surface roughness and estimate the frequency of rogue wave occurrence on the basis of the extremal theory for random fields. Statistical properties of wind-driven waves indicate to the fact that they may be described with a high precision by a homogeneous Gaussian random field of surface deviations from a mean level. Therefore, numerical models of homogeneous Gaussian fields appear to be efficient means for the sea surface simulation. To estimate the parameters of numerical models we made use of theoretical and real data. The results of the extremal theory for stochastic processes and fields allow us to calculate the mean number of extremely high waves above a given level for a known time interval according to the sea roughness model considered here.
This research was supported by the Russian Foundation for Basic Research (Project 18-31-00159).

11:50am - 12:10pm
Large Deviation Approximations for the Bagai Statistic
Presenter: Soshi KAWADA, Tokyo university of Science, Japan
Co-Authors: Hidetoshi MURAKAMI
Testing hypotheses is one of the most important challenges in performing nonparametric statistics. Various nonparametric statistics have been proposed and discussed for a long time. We use the exact critical value for testing hypotheses when the sample sizes are small. However, for large sample sizes, it is difficult to obtain the exact critical value. Under these circumstances, we must estimate the exact critical value with an approximation

method. In this report, we apply a strong large deviation approximations to distribution-free test statistic for stochastic ordering in the competing risks model. Then, we compare the approximations, saddlepoint approximation and normal approximation. We compute exact tail probabilities for small samples and relative errors of each approximations.

SL3: Statistical Learning - Methods and Applications III
Blue Lecture Hall (HS 402)

10:30am - 10:55am

Smart Recommendation System To Simplify Projecting For A Hmi/Scada Platform

Presenter: Kathrin PLANKENSTEINER, Fachhochschule Vorarlberg, Austria
Co-Authors: Sebastian MALIN, Robert MERZ, Reinhard MAYR, Sebastian SCHÖNDORFER, Mike THOMAS

In modern industry, HMI/SCADA software platforms are state-of-the-art for computer-aided monitoring and controlling of automated manufacturing processes. Representing a vast number of variables, linked to sensors and actuators from a variety of dierent machines in a uniform data model is an important feature of these systems. The current practice to manually enter the variables, each consisting of metadata like a unique name, data type, data source is considered time-consuming and expensive. Large automotive projects contain up to 1.6 million variables. This study presents an approach of a smart recommendation system simplifying the projecting of manufacturing plants for HMI/SCADA platforms. First experiments indicate that data of existing projects contain poor information for a reliable classication. Therefore, the paper discusses how semantic information can be added to improve the results. Data generated following standardized naming conventions, e.g., Weihenstephaner Standards, PackML, ANSI/ISA-95 is tested to achieve this goal. Thereby, a data model providing sucient information is developed such that not only classication, but even clustering, lead to an appropriate representation of the real manufacturing plant. Preliminary results indicate that using a standardized naming classication model can be used in a smart recommendation system.

10:55am - 11:20am

Microstructure Image Segmentation Using Patch-Based Clustering Approach

Presenter: Dženana ALAGIĆ, KAI Kompetenzzentrum Automobil- und Industrieelektronik GmbH, Austria
Co-Authors: Jürgen PILZ

In material science, the microstructure characteristics like the size and morphology of grains in the polycrystalline metal influence its physical and mechanical properties. Manually identifying and measuring features in microstructure images, such as Scanning Electron Microscopy (SEM) and Focused Ion Beam (FIB), is very slow, tedious and prone to errors. To overcome these issues, this work introduces a segmentation algorithm to automatically extract quantitative information from different types of microstructure images. Since labeled data is not provided, a patch-based clustering approach for image segmentation based on self-defined and textural features is proposed. The algorithm distinguishes

between two classes: grain and grain boundary area, making it effective on a variety of microstructure images. The performance of the fuzzy c-means algorithm and the Gaussian mixture model (GMM) is evaluated both visually and quantitatively using cluster internal and stability measures. The final, pixelwise segmentation is achieved with the Seeded Region Growing (SRG) algorithm using the identified grain areas as seed points.

11:20am - 11:45am

Estimation Of The Latent Signals For Consensus Across Multiple Ranked Lists Using Convex Optimisation

Presenter: Luca VITALE, Medical University of Graz, Austria
Co-Authors: Michael G. SCHIMEK

The ranking of the objects is widely used to assess the relative quality or relevance of such objects across the assessors. The assessors' decisions are independent and based on different metrics. The inconsistency of assessments is primarily due to the diversity of the involved metrics. We propose convex optimisation algorithms in combination with non-parametric Bootstrap for the estimation of the latent signals that inform the observed rankings. The goal is obtaining the optimal latent signals that permit to reduce the noise between the assessors. The order of the values of the estimated signals represents a consensus ranking of the relative importance of the objects across the observed individual rankings. The stability of each object can be evaluated by its standard error. The methodology is tested on simulated data to prove the real efficiency of the different algorithms.

11:45am - 12:10pm

Predicting Unusual Testing Results in Manufacturing Using Anomaly Detection

Presenter: Alicia MANGLANO, FH Vorarlberg, Austria
Co-Authors: Kathrin PLANKENSTEINER, Alberto VICECONTI

The ability to detect and consequently prevent anomalies during production is crucial and directly concerns the efficiency of industrial processes and goods. This study aims to thoroughly analyze and ultimately apply commonly used supervised and unsupervised techniques to a real-word industrial problem, in which products are tested prior to delivery. During production, each instrument is subjected to a variety of tests. If necessary, instrument parameters are calibrated during each of the testing stages. The data set currently available includes measurement values from one type of instrument over a three year time period. Despite one instrument has passed all test stages positively, it happens in rare cases that a final functional test is negative. Since each test is requiring significant resources and time, it is of utmost importance to predict the final test result as early as possible. This study focuses on the evaluation whether it is possible to forecast the final test result at any stage of the previously performed tests.

Since the industrial data generation process is complex and multi-sourced, several challenges arise when modeling the data. Although test bench related operations follow a

certain sequence, some instruments undergo the same operation in loops or sometimes, previous operations are repeated in order to stay cautious in the results. Consequently, the nature of the given data is partially non-independent-and-identically-distributed (non-iid), which forces a significant reduction on the number of observations.
Results indicate that it is possible to model the final measurement based on the previous test stages. It is shown, that interpolation performs well whereas extrapolation beyond the range might lack accuracy.

SSM2: Stochastic Modelling and Simulation in Materials Science and Engineering II
Auditorium Maximum (HS 401)

10:30am - 10:55am

Testing Goodness of Fit for Point Processes via Topological Data Analysis

Presenter: Christian HIRSCH, University of Mannheim, Germany
Co-Authors: Christophe BISCIO, Nicolas CHENAVIER, Anne Marie SVANE

In recent years, we are experiencing an explosion of the quantity, but above all the dimensionality and complexity of data. This represents a daunting challenge for classical statistical models. Topological data analysis aims to step up to these challenges by characterizing data via classical invariants from algebraic topology, most prominently the persistence diagram.
In this talk, we introduce tests for the goodness of fit of point patterns based on the persistence diagram. The tests rely on a functional central limit theorem for persistent Betti numbers induced by point processes with fast decay of correlations. We investigate the performance of these tests on simulated data and on a dataset from neuroscience.

10:55am - 11:20am

Homogenized Metrics in Planar, Multiscale Random Sets: Dilute Density of Grains

Presenter: François WILLOT, Mines ParisTech, France

The behavior of the length of geodesics (minimal paths) spanning planar, stationary random sets of grains, which are either rigid obstacles or pores, is investigated using numerical and theoretical bounds. This work is motivated by the classical problem in mechanics of porous and rigidly-reinforced perfectly-plastic media. The pseudometric is zero in pores, infinite in rigid obstacles, and equals the Euclidean distance in the embedding matrix. Accordingly, minimal paths tend to cross porous grains while avoiding rigid particles. Multiscale random structures based on Boolean sets of disk-shaped grains are considered, as well as random sequential adsorption models of squares, and we focus on the asymptotic limit of an infinitesimal density of grains. Our approach relies on numerical computations on the one hand, and on upper-bounds relevant to an infinitesimal grains density, on the other hand. Our results allow one to interpret certain predictions of nonlinear homogenization theories.

11:20am - 11:45am

Stochastic Modeling of Fiber-reinforced Ultra High Performance Concrete based on 3D Image Analysis

Presenter: Konstantin HAUCH, Technische Universität Kaiserslautern, Germany
Co-Authors: Kasem MARYAMH, Claudia REDENBACH, Jürgen SCHNELL

Desirable properties of concrete in buildings are high compressive and tensile strength and load-bearing capacities. Ultra High Performance Concrete (UHPC) is characterized by a high compressive strength. The load-bearing capacity and the tensile strength can be increased by adding steel fibers to the concrete. The orientations and the positions of the fibers in the concrete considerably determine its properties. For instance, the load-bearing capacity is maximal if the direction of the loading force is perpendicular to the fiber orientation. A homogeneous distribution of the fibers in the concrete leads to a homogeneous load-bearing capacity. The orientations and positions of the steel fibers depend on many parameters in the production process. Length, diameter and volume fraction of the fibers, the amount of superplasticizer, the pouring direction, and the formwork for the concrete are just a few examples of these parameters. To investigate these dependencies, a large number of samples of fiber-reinforced UHPC with varying parameters was produced. The samples were imaged by using micro computed tomography and the fiber system was reconstructed by a suitable segmentation algorithm. Statistical analysis showed that the fiber diameter and the amount of superplasticizer have an effect on the orientations and the spatial distribution of the fibers. Furthermore, the orientation of the fibers changes depending on their position in the UHPC. In bending tests, the load-bearing capacity of the fiber-reinforced UHCP is measured. CT imaging additonally allows for an investigation of correlation between crack locations and the local fiber geometry. Finally, a method for predicting the mechanical properties of the fiber-reinforced UHPC using a stochastic model is outlined.

11:45am - 12:10pm

Stochastic 3D Microstructure Modeling of Differently Compacted Cathodes in Lithium-ion Batteries

Presenter: Benedikt PRIFLING, Institute of Stochastics, Ulm University, 89069 Ulm, Germany
Co-Authors: Daniel WESTHOFF, Denny SCHMIDT, Henning MARKÖTTER, Ingo MANKE, Volker KNOBLAUCH, Volker SCHMIDT

Due to their preferable electrochemical properties, lithium-ion batteries are widely used as energy storage devices in a wide range of applications. It is known that the electrode morphology strongly influences the overall battery performance. Therefore, a deeper understanding of the microstructure is an important task regarding the optimization of the electro-chemical performance. Thus, the detailed investigation of the manufacturing process and its impact on the microstructure is an essential part of battery research. The 3D morphology of these differently compacted cathodes has been obtained by synchrotron tomography. Using the results of a comprehensive statistical analysis, we are able to calibrate one and the same type of a parametric stochastic 3D microstructure model to all

compaction loads. This allows us to describe the complex 3D morphology of the cathode by a small number of parameters. By least-squares regression analysis, we are finally able to predict the model parameters and hence the morphology of cathodes for arbitrary compaction loads. This can be used to provide a wide spectrum of virtual but realistic 3D microstructures as valuable input for numerical simulations of charge transport.

OD2: Optimal Design in Mixed Models
Green Lecture Hall (HS 403)

1:00pm - 1:25pm

Optimal Design for Multi-stage Experiments

Presenter: Emily Sarah MATTHEWS, University of Southampton, United Kingdom
Co-Authors: David Christopher WOODS

Experiments in which the same experimental unit has factors applied to it in different stages are common in industry. In this talk, we discuss approaches to finding optimal designs for multi-stage experiments, including a pseudo-Bayesian compound objective function. We present multi-stage designs where mixed effect models are assumed to model the response from at least one stage of the experiment. Our work is demonstrated through an example from the pharmaceutical industry.

1:25pm - 1:50pm

On the Construction of D-optimal Designs in Mixed Binary Regression Models

Presenter: Parisa PARSAMARAM, Otto von Guericke University, Germany
Co-Authors: Rainer SCHWABE

Mixed effects regression models have recently been utilized in a wide variety of applications. Some of these applications contain a binary response variable such that mixed binary regression models become appropriate. In the present work the two link functions of the logit and the probit link are considered which result in the mixed logistic and mixed probit regression models, respectively. The aim is to propose D-optimal designs for these models. In particular, those designs are obtained by maximization of the D-optimality criterion which is defined as the determinant of the Fisher information matrix. As for the present models under consideration the Fisher information cannot be determined explicitly in a closed form expression, it becomes necessary to propagate an analytical approximation which is as precise as possible. One suggestion is to derive the quasi Fisher information matrix based on generalized estimation equations. However, also this approach does not lead to an explicit closed form. To circumvent this problem a reliable analytical approximation is recommended which fulfills the needs for attaining D-optimal designs. As examples two particular cases are considered where either one or two random effects are assumed to present, and approximately locally D-optimal designs are obtained. Subsequently, the so obtained designs are checked by means of the equivalence theorem, and their reliability is assessed by a simulation study on the finite sample size behavior of the mean squared error and a comparison with numerically D-optimal designs based on numerical computations.

1:50pm - 2:15pm

Standardized Maximin Optimal Designs For Mixed-Effects Poisson Regression Models

Presenter: Marius SCHMIDT, Otto-von-Guericke-Universität Magdeburg, Germany
Co-Authors: Rainer SCHWABE

The Poisson regression model is frequently used for modeling count data, which arises in experiments where the number of objects or occurrences of events of interest is observed. In such experiments there may be repeated measurements for each statistical unit. Assuming a Gamma distributed block effect for each statistical unit, we obtain the Poisson-Gamma model, which is a nonlinear mixed-effects regression model and a generalization of the Poisson model.

The information matrix for the Poisson-Gamma model, which can be represented in terms of the information matrix for the Poisson model, is derived analytically. Optimal designs are based on the optimization of a function of the information matrix, for example, for D-optimality the determinant is maximized. Since the Poisson-Gamma model is nonlinear, the optimal designs depend on the unknown parameters. Thus parameter misspecification can lead to poor designs. To obtain more robust designs we derive standardized maximin D- and c-optimal designs, which maximize the worst efficiency with respect to a prespecified parameter set.

2:15pm - 2:40pm

Efficient Sample Sizes For Trials With Clustered Data: How To Deal With Unknown And Heterogeneous Variance Parameters, And With Cluster Size Variation, In The Design Stage ?

Presenter: Gerard VAN BREUKELEN, Maastricht University, Netherlands
Co-Authors: Math CANDEL

The sample size needed for a randomized trial depends on the outcome variance, which is never known in advance. This problem is aggravated in a cluster randomized trial (CRT), which randomizes organizations (clusters) instead of individuals to treatment, a design often used in public health and education. In a CRT the optimal sample size must be determined at two levels (number of clusters, number of persons per cluster), and it depends on the outcome variance between and within clusters, and on the study cost per cluster resp. person, which can all be treatment-dependent. The optimal design of a two-arm CRT with quantitative outcome thus depends on four unknown variances, one per level per arm. One solution is Maximin design (MMD), which maximizes the minimum efficiency of the design over the parameter and design space. We present a closed form solution for the MMD and compare it with the balanced design (Statistics in Medicine, 2018). Further, we present a simple sample size adjustment to compensate for the efficiency loss incurred by unplanned sample size variation between clusters within the same treatment arm. This adjustment only requires specification of the coefficient of variaton of sample size between clusters, and will be seen to perform quite well for a range of realistic cluster size distributions (Statistics in Medicine, 2007, 2016).

2:40pm - 3:05pm

Optimal Designs in Mixed Effects Multi-Factor Models

Presenter: Maryna PRUS, Otto von Guericke University Magdeburg, Germany

Optimal designs in multi-factor models without random effects have been well discussed in the literature. It has been established that there is a structural relationship between the information matrices in multi- factor models and the corresponding univariate marginal models. For some particular cases optimal designs in the multi-factor models can be constructed as product-type designs. However, these results cannot be directly transferred to mixed effects models, because the resulting design criteria do not allow for a direct factorization. We propose analytical solutions for optimal designs in mixed effects multi-factor models with particular covariance structures of the random effects and illustrate the results by simple examples.

OSP1: Optimal Selection Procedures Blue Lecture Hall (HS 402)

1:00pm - 1:25pm

Introduction to Selection Procedures and their Robustness

Presenter: Dieter RASCH, Boku Wien, Austria

The Bechhofer and the Gupta approach for selecting the best out of a given distributions are described. Especially the best distribution is that of a given normal distributions with the largest expectation. It is shown that an optimal combination of both approaches leads to a smaller overall sample size than Bechhofers approach alone and can be recommended for general use.

Reference: Rasch, D. and Schott, D. (2018) Mathematical Statistics, Wiley, Oxford.

1:25pm - 1:50pm

The Calculation of the Coefficients of the Fleishman Distributions

Presenter: Hans Dieter SCHOTT, Hochschule Wismar, Germany

The Fleishman system of distributions is often used in simulation experiments to generate non-normal distributions with given skewness and kurtosis. For each pair (of skewness, kurtosis) a member of the system exists. A Fleishman distribution is generated from a standard normal random variable x (with mean 0 and variance 1) by

$$y = a + bx + cx^2 + dx^3$$

with real parameters a, b, c, d.

In the paper it is discussed how for any admissible pair (of skewness, kurtosis) the coefficient a, b, c, d can be calculated from a nonlinear system of algebraic equations using computer algebra. The general solvability of this system is considered. Although in the references only solutions with positive coefficients b, c, d occur, the properties of distributions connected with the other real solutions are investigated, too.

References:

Fleishman, A.J. (1978): A method for simulating non-normal distributions. Psychometrika43: 521–532.
Rasch, D.: Mathematische Statistik, 217–218. Johann Ambrosius Barth Verlag, Leipzig 1995.

1:50pm - 2:15pm
Simulating Non-normal Distributions and the Robustness of the Bechhofer Selection
Presenter: Philipp SCHNEIDER, Alpen-Adria-Universität Klagenfurt, Austria
References: Rasch, D. and Yanagida, T. (2019). An optimal two-stage procedure to select the best out of a Normal population. J Stat Theory Pract.13, 1.

2:15pm - 2:40pm
How good is a Two-Stage Approach for Non-normal Distributions?
Presenter: Dieter RASCH, Boku Wien, Germany
Co-Authors: Schneider PHILIPP
By a large simulation experiment using ten different (non-normal) Fleishman distributions with different (skewness, kurtosis)-pairs we could show that only for two extreme distributions no 80% - ne robustness of the two-stage approach coukd be found. Most of the other distributions arrived a 90% robustness.

2:40pm - 3:05pm
The Two-Stage Selection Procedure for $t > 1$
Presenter: Jürgen PILZ, Alpen-Adria-Universität Klagenfurt, Austria
In this paper the reults of Rasch and Yanagida (2019) are generalised for the case that more than one best population has to be selected.
Reference: Rasch, D. and Yanagida, T. (2019). An optimal two-stage procedure to select the best out of a Normal population. J Stat Theory Pract.13, 1.

AN1: Analytical and Numerical Methods in Statistics
Auditorium Maximum (HS 401)

3:05pm - 3:30pm
Monotone Functions Generated By Improper Parameter Integrals And Their Significance For Statistical Inequality Problems
Presenter: Hans Dieter SCHOTT, Hochschule Wismar, Germany
We investigate the analytical and numerical solution of inequalities established by monotone functions which are generated by improper integrals. We present some important applications concerning cumulative distribution functions as determination of quantils or of sample sizes in statistical selection problems. The starting point is the Bechhofer selection problem [1] which is also discussed in [2] and [3].
Reference:

[1] Bechhofer, R.E.: A Single Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. Ann. Math. Statist. 25,16-39 (1954).
[2] Rasch, D. and Schott, D.: Mathematical Statistics. Wiley 2018.
[3] Schott, D.: Some remarks on a Statistical Selection Procedure of Bechhofer for Expectations. Rostock. Math. Kolloq. 71, 57-67 (2017/18).

3:30pm - 3:55pm

On a Parametrical Estimation for a Convolution of Exponential Densities

Presenter: Nadezda SPIRIDOVSKA, Transport and Telecommunication Institute, Lomonosov Str., 1, LV-1019, Riga, Latvia
Co-Authors: Alexander ANDRONOV, Diana SANTALOVA

Broad application of the continuous time Markov chain is caused by memoryless property of exponential distribution. An employment of non-exponential distributions leads to remarkable analytical difficulties. The usage of arbitrary nonnegative density approximation by a convolution of exponential densities is a way of considerable interest.
Two aspects of the problem solution are considered. Firstly, the parametrical estimation of the convolution on the basis of given statistical data. Secondly, an approximation of fixed non-negative density.
An approximation and estimation are performed by the method of the moments, maximum likelihood method, and fitting of a density. An empirical analysis of different approaches has been performed with the use of simulation.
The efficiency of the considered approach is illustrated by the task of the queuing theory.

3:55pm - 4:20pm

Numerical functional kernel Monte Carlo algorithm

Presenter: Anton Vaclavovich VOYTISHEK, Novosibirsk State University, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Russian Federation
Co-Authors: Tatyana Evgenyevna BULGAKOVA

This talk continues our investigations on numerical (implemented on a computer) randomized functional algorithms for approximation of a solution of Fredholm integral equation of the second kind (see, for example, the talk: Shipilov N. M., Voytishek A. V. On conditional optimization of the randomized projection and projection-mesh functional algorithms for numerical solution of the Fredholm integral equations of the second kind // Book of abstracts of the Ninth International Workshop on Simulation. Barselona, Spain: Polytechnic University of Catalonia. 2018. P. 119). The kernel projection-mesh algorithm is one of the most effective method for numerical solution of practically important problems. The recomendations on choice of parameters and special functions for this algorithm are presented.

ESA2: Ergodicity and sensitivity analysis of stochastic systems and networks II
Blue Lecture Hall (HS 402)

3:05pm - 3:30pm

About Ergodicity And Convergence Rate Of A Generalized Reliability System With Warm Reserve

Presenter: Galina ZVERKINA, RUT (MIIT), ICS RAS, Russian Federation

The behaviour of a generalized reliability system with warm reserve is described by non-regenerative process. We consider the case when the intensity of failure and restoration of both elements of the system are between constant or variable bounds. By natural Markovisation of this process, and using coupling method, we prove the ergodicity of this process. The rate of convergence of the distribution of Markovisated process to the invariant measure depends on the type of restrictions for the intensities. It can be exponential or polynomial. We give the algorithm of the construction of the strong bounds of convergence of the distribution of Markovisated process in total variation metrics. Therefore, the bounds of convergence of original process are the same.

3:30pm - 3:55pm

Lyapunov Function Application To The Proof Of The Recurrence Of Some Queueing Process

Presenter: Alexander Yu. VERETENNIKOV, University of Leeds, UK; National Research University Higher School of Economics, and Institute for Information Transmission Problems, United Kingdom

Co-Authors: Galina A. ZVERKINA

Two dependent queuing systems that may be in working or non-working states are considered. The intensities of failures and recoveries of both queuing systems and the values inverse to them are assumed to be bounded above by linear functions with some conditions on the coefficients. The joint maintenance process is not regenerative, and its ergodicity is not obvious.

3:55pm - 4:20pm

Sensitivity Analysis Of k-out-of-n System Characteristics To Shapes Of Their Components Life And Repair Times Distributions

Presenter: Vladimir Vassilievich RYKOV, Gubkin Russian State University of Oil and Gasd, Russian Federation

Co-Authors: N. IVANOVA

The paper deals with the problem of k-out-of-n system reliability characteristics sensitivity to the shape of their elements life and repair times distribution under ”quick” restoration. This research eld has wide application in various elds, so its study dates back to the middle of the last century. In this area, a signicant contribution was made by V.A. Sevastyanov, B.V. Gnedenko, I.N. Kovalenko, A.D. Soloviev and V.V. Rykov. In recent works, a hot redundant system is considered with exponential distribution of its elements life and repair time as well as general distribution of their repair time. The results show their asymptotic insensitivity to the shape life and repair time distributions with the help of simulation method. Sensitivity analysis of k-out-of-n system was carried out with the help of simulation modeling, the results of which prove the system insensitivity to the

shape of distribution function with "quick" restoration.

SNM1: Stochastic numerical methods Green Lecture Hall (HS 403)

3:05pm - 3:30pm

Random Search Method with a "Memory" for Global Extremum of a Function

Presenter: Liudmila Vasilevna VLADIMIROVA, St. Petersburg, Russian Federation
Co-Authors: Sergey Michaylovich ERMAKOV

The general scheme of stochastic global optimization methods can be represented as follows. In the region D of extremum search for the function $f(X)$, N points $X_j (j = 1, \ldots, N)$ are chosen randomly or quasi-randomly and N values $f(X_j)$ are calculated. Of the N points, m points are stored, where f values are the largest (smallest). The set of these m points is called the zero generation. After this, the iterative Markov algorithm is executed. If the k-th generation of m_k points is determined, the method is specified to of obtain the $(k+1)$-th generation of m_{k+1} points. The methods mentioned provide the sequence of generations to converge with probability 1 to the global extremum point.
Our report discusses one of methods of this kind proposed by the authors in 1977.
The proposed method idea is to construct the normal density on the basis of k-th generation points. The points of the next generation are sampled from the normal distribution. The number of points decreases with k growth. On final stages it is advisable to use the gradient method.
Random extremum search with covariance matrix (search with "memory") is convenient for solving problems of charged beam dynamics optimization. Such problems are dedicated to minimization of quality functional by control parameters.

3:30pm - 3:55pm

Symptom Analysis of Multidimensional Categorical Data with Application in Genetics

Presenter: Evgeniia Petrovna SKURAT, St. Petersburg State University, Russian Federation
Co-Authors: Nina Petrovna ALEKSEYEVA, Fatema Sadik AL-JUBOORI

In statistical analysis of multidimensional categorical data, there is a problem of the dimension reduction which means search a few functions of factors with the least information loss. When analysis of factors separately or linear statistical methods are not enough to describe the risk group, then we can use the symptom-syndrome model in which predicate was expressed in terms of independent factors as linear combinations over eld F_2, which form the nite projective space and called symptoms for short. The Gehan Wilcoxon test was used in survival symptom analysis. If we construct a nite projective space not for k dichotomous variables, but for their $2^k - 1$ various non-degenerate multiplications without repetition, then we receive Zhegalkin polynomials which describe all sorts of logical functions. Using the parameterization of all possible logical combinations by means Zhegalkin polynomials we can highlight the risk group and receive the most informative combination of factors for survival analysis.

3:55pm - 4:20pm

The Monte Carlo Method for a Solution of ODE System

Presenter: Tatiana Mikhailovna TOVSTIK, Sankt-Petersburg University, Russian Federation

Co-Authors: Sergei Mikhailovich ERMAKOV

The Monte Carlo method is an effective method for a solution of large dimension problems. An elaboration of methods of solution of Cauchu problems for large systems of ODE based on a simulation of Markov processes is very interesting, but this problem is not enough investigated now.

Here this problem is discussed and some peculiarities are indicated. As a rule differential equations are replaced by equivalent integral Volterra equations, and then the well known Neumann-Ulam system is used (see Ermakov S.M., 2009).

The appearing peculiarities are discussed. First, the linearization that complicate the Markov process construction is to be used. The case of a polynomial nonlinearity is an exclusive one, because there exist algorithms based on the approximation of nonlinear functions by polynomials and on branching processes simulation. In the general case the interval $[0, t]$ is divided by subintervals of length and the approximation is performed in each subinterval. Two types of errors, systematic and random, appear and they are to be investigated.

Second, in the linear case the value t may be arbitrary large, but in the nonlinear case the choice of depends as on the small value of error of polynomial approximation, so on the value of the Picard interval of the solution existence.

As examples, some linear and nonlinear systems ODE by the Monte Carlo method are solved. The simulation of branching Markov chains is used. The obtained solutions are compared with the solutions found by the Runge-Kutta method.

The error of the proposed methods is to be investigated in future, but the considered examples point of their perspective.

The work is supported by RFBR, №17-01-00267-a.

Poster Session

Tuesday September 3, 2019

PS: Student Poster Session
Foyer Blue Lecture Hall (HS 402)

1:00pm - 1:30pm

A Class of Score Functions for the Analysis Kinematic Handwriting Data

Presenter: Cami Marie FUGLSBY, South Dakota State University, United States of America

Co-Authors: Christopher SAUNDERS, Danica OMMEN2, Michael CALIGIURI

Recent concerns about the validity of forensic handwriting analysis have led to a number of research projects focused on the relationship between how a writer writes (i.e. kinematic features of handwriting samples) and writership determinations made by examiners. One component of these studies is to be able to measure the dissimilarity of a pair of handwriting samples based on the recorded kinematic features. The kinematic features are constructed by decomposing a writing sample into the individual strokes of the handwritten phrase. In this presentation we will present a class of dissimilarity scores constructed for measuring the difference between the multivariate distributions of two sets of strokes based on quantile goodness of fit comparisons. The score functions were evaluated using receiver operating characteristic (ROC) curves constructed from all pairwise comparisons of writing samples.

Reproducibility of High-Resolution 3D Bullet Scans and Automated Bullet Matching Scores

Presenter: Kiegan E RICE, Iowa State University, United States of America
Co-Authors: Heike HOFMANN, Ulrike GENSCHEL

Recent critiques of forensic firearms comparisons have led to an increase in research focused on automated bullet matching algorithms. These automated approaches use high-resolution 3D scans of bullet land engraved areas (LEAs), with the goal of increasing objectivity and reducing human involvement in the comparison process. Translating physical LEAs to high-resolution 3D scans introduces human microscope operators into the process. To address the potential impact of human involvement in the process, we designed a repeatability and reproducibility (R&R) study to quantify sources of variability. Repeated scans were captured using multiple operators, microscopes, and bullets. We will present results from this study, quantifying the observed differences on extracted LEA signatures as well as the impact on accuracy of an automated bullet matching algorithm. We will also discuss the sources of variability and which elements of data collection have the largest impact on algorithm performance.

A Bayesian Hierarchical Model for Forensic Writer Identification

Presenter: Amy CRAWFORD, Iowa State University, United States of America
Co-Authors: Nicholas BERRY, Alicia CARRIQUIRY, Danica OMMEN

We provide a method for performing automated writer recognition within a closed set of writers. We begin by extracting data from scanned handwritten documents using a new automated, open source process. The result is a collection of handwriting components that are roughly the size of letters. We call these components glyphs, and treat them as small graph structures with vertices and edges. We demonstrate how writers can be suitably characterized by the rate at which they emit glyphs to certain classes with their writing. We assign glyphs into such classes using a clustering algorithm with a novel distance metric. We also make use of a deterministic method to create glyph classes, and results from the two methods are compared. Glyph class memberships serve as data for a Bayesian hierarchical model with a multinomial data distribution. As is often done with

count data, we evaluate a measure of under-dispersion for the model, and show how the dispersion is meaningful to the writer identification application beyond its traditional use as a model evaluation tool. Samples from 27 writers in the Computer Vision Lab database are used in a full scale writer identification.

Interval-Wise Testing Of Functional Data Defined On Two-Dimensional Domains

Presenter: Patrick B. LANGTHALER, Paris-Lodron University Salzburg, Austria
Co-Authors: Alessia PINI, Arne C. BATHKE

In statistical practice one can find oneself confronted with data that has been measured on a fine two-dimensional grid. Examples are geographical data (e.g. temperature, precipitation, wind speed) or medical data like High Density Electroencephalography (HD-EEG) measurements. Usually this grid represents the discretization of some underlying smooth function. The data is therefore often characterized as functional. A possible research hypothesis is the existence of a difference in mean values of the outcome variable in some subset of the grid between two or more experimental groups. A common problem is, that this subset cannot be specified a priori but discovering on what subset the outcome differs is indeed part of the research question. In this case a control of the Family Wise Error Rate (FWER) has to take place. For high-dimensional data, classical methods that control the FWER in the strong sense (e.g. Bonferroni, Bonferroni-Holm, Closed Testing Procedure) can be very conservative, leaving little statistical power to discover potential effects of interest. We therefore discuss a method for controlling the FWER on certain neighbourhoods (e.g. squares) contained within the grid. The resulting control of the FWER is more flexible and weaker than classical methods, resulting in more statistical power, while still allowing to declare significant differences on certain neighbourhoods after testing.

Single-neuron Representations of Ambiguous Words in the Human Temporal Lobe

Presenter: Bita SAMIMIZAD, University of Bonn, Germany
Co-Authors: Thomas P. REBER, Jan BOSTRÖM, Christian E. ELGER, Florian MORMANN

Due to their ambiguity, homonyms are ideal for researching the semantic representation in concept cells, i.e., units with semantically invariant responses to a specific concept. Homonyms are phonologically and orthographically identical words that represent more than one concept. Using homonyms as stimulus material for our study, we addressed questions such as: Given a response-eliciting homonym (identified by a binwise Wilcoxon rank-sum test), will the neuron also respond to the corresponding meanings and if so, will we observe the activation of both or only one of the meanings (multiple access or unitary perception)? Furthermore, will providing a biasing context modulate the firing rate of the neuron? Our results show that most units responded to only one corresponding meaning rather than both. Moreover, we observed a prolonged response to the homonym if it is

presented in the context of the preferred meaning. Our findings thus provide evidence in favor of the multiple-access theory and against the unitary-perception theory.

Altered Behavior of Epileptic Interneurons and Principal Cells during Cognitive Processing

Presenter: Gert DEHNEN, Dept. of Epileptology, University of Bonn Medical Center
Co-Authors: Jan BOSTRÖM, Christian E. ELGER, Florian MORMANN

Previous studies have shown that single units in the human medial temporal lobe (MTL) can encode sparse, selective and invariant representations of the environment. We looked for differences in neuronal processing during object recognition specifically related to the epileptic process, in different MTL regions, by comparing response characteristics of single neurons and spike timing relative to ongoing oscillations in the local field potential (LFP) between the epileptic and the contralateral hemisphere.

The most striking difference between the epileptic and contralateral hemisphere was the fraction of responsive units (Fishers-exact-test: $p = 2.5 \cdot 10 - 16$). Analyzed dynamic properties such as firing rate, peak and onset latency in response to stimulus presentation largely showed the same behavior. Phase locking of action potentials to ongoing oscillations, determined via Hilbert transform and Rayleigh test, showed a significant difference between the epileptic and contralateral hemisphere, particularly in the theta and delta band.

We hypothesize that reduced spike timing variability may cause the reduced neuronal responsiveness in the epileptic hemisphere.

A Nonparametric Approach of Interpreting the Rejection Step in the ABC-Algorithm

Presenter: Victoria RACHER, University of Salzburg, Austria

The likelihood plays a central role in model-based statistical inference as it expresses the probability of observed data given a certain statistical model. However, in complex models, it is often impossible to derive a closed form of the likelihood or its optimum. Or, from a computational point of view, the likelihood function might be very costly to evaluate.

One method to handle this challenge is Approximate Bayesian Computing (ABC). In a nutshell, the ABC-Algorithm approximates the true but unknown likelihood function by simulation following a simple trial and error scheme. The key step in the ABC-Algorithm is the comparison of two datasets.

The ABC-Algorithm does not require any assumptions, but this has the consequence that the algorithm contains somewhat arbitary components which in turn leads to difficulties regarding the compatibility of results.

Starting with the Basic ABC-Algorithm, we provide a first nonparametric approach in order to reduce the arbitary components. Specifically, we consider the classical nonparametric relative effect introduced by Mann and Withney, which can be interpreted as a nonparametric location-measure, as well as the overlap index introduced by Parkinson et.

al in 2018, which is a nonparametric dispersion-measure. Based on the relative effect and the overlap index, we are able to construct a multiple hypothesis test and thus interpret the classical ABC-Algorithm as a decision problem. Simulations indicate very promising results.

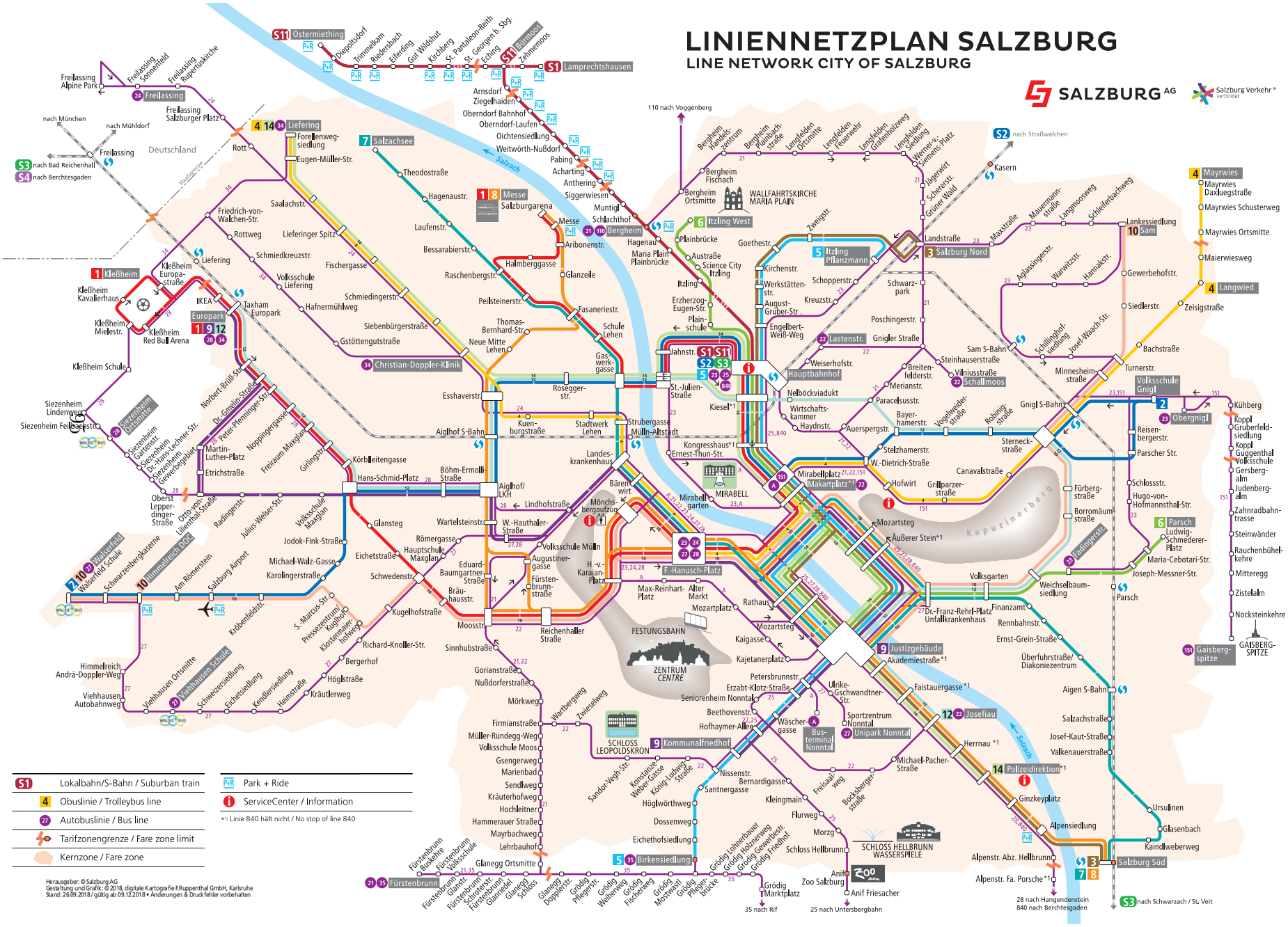
LINIENNETZPLAN SALZBURG

LINE NETWORK CITY OF SALZBURG

SALZBURG AG

Salzburg Verkehr * verbindet

Bus-lines



S1 Lokalbahn/S-Bahn / Suburban train

4 Obuslinie / Trolleybus line

27 Autobuslinie / Bus line

Tarifzonengrenze / Fare zone limit

Kernzone / Fare zone

P+R Park + Ride

ServiceCenter / Information

*1 Linie 840 hält nicht / No stop of line 840

Herausgeber: © Salzburg AG
Gestaltung und Grafik: © 2018, digitale Kartografie F.Ruppenthal GmbH, Karlsruhe
Stand: 26.09.2018 / gültig ab 09.12.2018 • Änderungen & Druckfehler vorbehalten